Classical tests

Science unit WCS Indonesia

5-9th of August 2019

Quiz 1

- A parametric test assumes:
  - no distribution of observed data
  - quantiles of observed data
  - standard error of observed data
  - normal distribution of observed data

Answer Quiz 1

- A parametric test assumes:
  - no distribution of observed data
  - quantiles of observed data
  - standard error of observed data
  - **normal distribution of observed data**

Quiz 2

- If you want to know whether a female lion runs faster than male lion, and you know that the data collected is normally distributed, you would use:
  - Kruskall Wallis test
  - t-test
  - ANCOVA test
  - chi-squared test

Answer Quiz 2

- If you want to know whether a female lion runs faster than male lion, and you know that the data collected is normally distributed, you would use:
    - Kruskall Wallis test
    - **t-test**
    - ANCOVA test
    - chi-squared test

Quiz 3

- If you want to know whether a grouper fish caught in Sigli, Tj Luar, and Kepulauan Seribu are different in body length, and you do not make any assumption of the data distribution, you would use:
  - Kruskall Wallis test
  - t-test
  - ANCOVA test
  - chi-squared test

Answer Quiz 3

- If you want to know whether a grouper fish caught in Sigli, Tj Luar, and Kepulauan Seribu are different in body length, and you do not make any assumption of the data distribution, you would use:
    - **Kruskall Wallis test**
    - t-test
    - ANCOVA test
    - chi-squared test

Quiz 4

- If you want to know whether a lovebird is sold due to its different sizes and or due to the area the shop is located, and you assume normal distribution of the observed data you would use:
    - Kruskall Wallis test
    - t-test
    - ANCOVA test
    - chi-squared test

Answer Quiz 4

- If you want to know whether a lovebird is sold due to its different sizes and or due to the area the shop is located, and you assume normal distribution of the observed data you would use:
  - Kruskall Wallis test
  - t-test
  - **ANCOVA test**
  - chi-squared test

Quiz 5

- If you want to know whether a survey on whether female or male eat more sharks depends on the education level of female/male, then we could use:
  - Kruskall Wallis test
  - t-test
  - ANCOVA test
  - chi-squared test

Answer Quiz 5

- If you want to know whether a survey on whether female or male eat more sharks depends on the education level of female/male, then we could use:
  - Kruskall Wallis test
  - t-test
  - ANCOVA test
  - **chi-squared test**

Classical tests

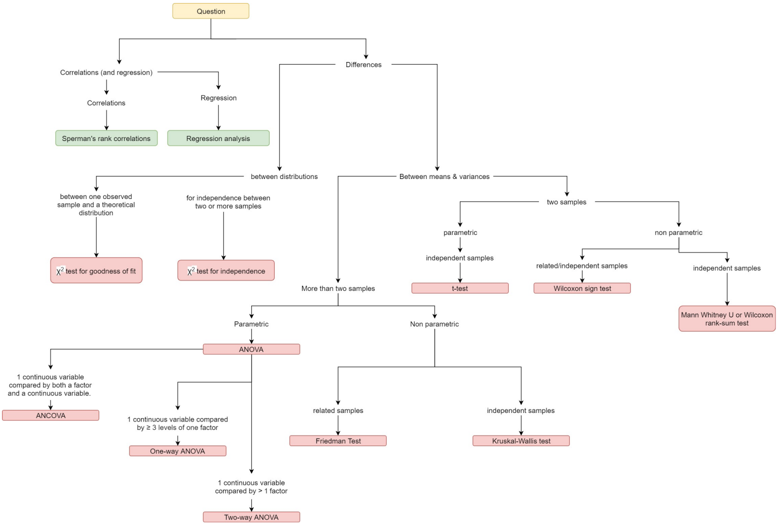


Figure 1: Classical tests diagram

Classical tests: Parametric vs non- parametric

- A parametric test assumes normal distribution of observed variables.
- A non-parametric test assumes no distribution of observed variables.

Parametric test

`t-test`

- `t-test` assumes normal distribution.
- **One-sample t-test** is used to compare a known mean value with a certain distribution.

t-test case study

We assume the daily energy intake in kJ for 11 gorillas to be:

```
daily.intake<-c(5260,5470,5640,6180,6390,6515,
6805,7515,7515,8230,8770)
summary(daily.intake) #summary statistics
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    5260    5910    6515    6754    7515    8770
```

Figure 2: Gorilla

t-test case study

Is the intake value deviates from a recommended value of 7725 kJ?

```
t.test(daily.intake,mu=7725)
```

```
##
##  One Sample t-test
##
## data:  daily.intake
## t = -2.8208, df = 10, p-value = 0.01814
## alternative hypothesis: true mean is not equal to 7725
## 95 percent confidence interval:
##  5986.348 7520.925
## sample estimates:
## mean of x
##  6753.636
```

Conclusion: mean daily intake is not equal to the recommended value (H_0 is rejected).

Two-sample t-test

- A **two-sample t-test** tests: if two samples mean ($barx_1$ and $\bar{x}_2$) may come from the same distribution or not ($\mu_1 = \mu_2$).
- We can then calculate:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SEDM}$$

- Where the **standard error of difference of means** is:

$$SEDM = \sqrt{SEM_1^2 + SEM_2^2}$$

.

Two-sample t-test case study

We want to calculate whether there are differences between the energy expend by obese gorilla vs lean gorilla.

```
library(ISwR)
attach(energy)
head(energy,4)
```

```
##   expend stature
## 1   9.21   obese
## 2   7.53    lean
## 3   7.48    lean
## 4   8.08    lean
```

```
t.test(expend~stature,var.equal=T) #expend (energy spent) is described by stature (height)
```

```
##
##  Two Sample t-test
##
## data:  expend by stature
## t = -3.9456, df = 20, p-value = 0.000799
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.411451 -1.051796
## sample estimates:
##  mean in group lean mean in group obese
##            8.066154           10.297778
```

```
                                   #we assume the variance is the same in two groups
detach(energy)
```

Two-sample t-test case study

- We found that the H_0 is rejected (p-value <0.05) and the alternative hypothesis is accepted.
- Conclusion: energy spent by obese vs lean gorillas are different.

Paired t-test

- **Paired t-test**: two measurements on the **same** experimental unit.
- **Pre- and post-** measurements/experiments.

Paired t-test case study

- Difference between the energy intake pre and post-release into the wild (after captured by hunters) in a group of gorillas.

```
attach(intake)
head(intake,3)
```

```
##    pre post
## 1 5260 3910
## 2 5470 4220
## 3 5640 3885
```

```
t.test(pre,post, paired=T)
```

```
##
##  Paired t-test
##
## data:  pre and post
## t = 11.941, df = 10, p-value = 3.059e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1074.072 1566.838
## sample estimates:
## mean of the differences
##                1320.455
```

```
    #the paired=T is important because it indicates that the sample
    #comes from the same experiment
detach(intake)
```

Paired t-test case study

- There is a significant difference between the energy intake before and after release into the wild (p-value <0.05).

Analysis of variance (ANOVA)

One-way ANOVA

- With more than two groups: use ANOVA (parametric).
- **One-way analysis of variance** : **total variation** (SSD_T) is a sum of **variaton within groups** (SSD_W) and **variation between groups** (SSD_B).
- Total variation is split into: describing **between group means** and between **individual measurements** within the groups.

One-way ANOVA

- The sum of the two variations can only be positive
- If there is no systematic differences **between the groups**, the sum of squares should be **partitioned** by **degrees of freedom**
  - $k - 1$ for SSD_B ($k =$ number of groups)
  - $n - k$ for SSD_W ($N =$ total number of observations).

One-way ANOVA

- **Mean squares** or pooled variance:

$$MS_W = SSD_W/(N-k)$$

$$MS_B = SSD_B/(k-1)$$

- With no **true group effect**, MS_B will be an estimate of σ^2
- If there is a group effect, MS_B will tend to be large.

One-way ANOVA

- So, a test for differences between a two group **means** can be done by comparing the **variance** estimates.
- The approach uses F distribution:

$$F = MS_B/MS_W$$

- F is ideally 1, but some variation around the value is expected.

One-way ANOVA case study

- Simple analyses of variance can be performed using function `lm` (similar to regression analysis).
- Example data of "red cell folate" (Altman 1991, p. 208). But assume the data is about the number of Pangolin scales confiscated in year 2013, 2014, and 2015.

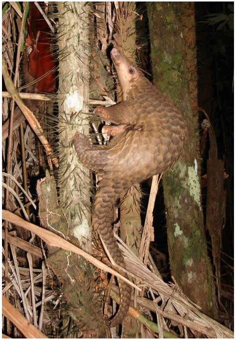


Figure 3: Pangolin

One-way ANOVA case study

- One-way analysis of variance is analogous to regression analysis
- But the **explanatory variables** are **factors** not numeric variable.

One-way ANOVA case study

```
head(pangolin1,3)
```

```
##   scales     year
## 1    243 year2013
## 2    251 year2013
## 3    275 year2013
```

```
anova(lm(pangolin1$scales~pangolin1$year))
```

```
## Analysis of Variance Table
##
## Response: pangolin1$scales
##                Df Sum Sq Mean Sq F value  Pr(>F)
## pangolin1$year  2  15516  7757.9  3.7113 0.04359 *
## Residuals      19  39716  2090.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- We get SSD_B and MS_B (top line)
- SSD_W and MS_W (second line).
- If the F test shows difference between groups (p=value $<$0.05), we need to know where the difference lies.

One-way ANOVA case study

- We should then compare the individual groups:

```
summary(lm(pangolin1$scales~pangolin1$year))
```

```
##
## Call:
## lm(formula = pangolin1$scales ~ pangolin1$year)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -73.625 -35.361  -4.444  35.625  75.375
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)              316.62      16.16  19.588 4.65e-14 ***
## pangolin1$yearyear2014   -60.18      22.22  -2.709   0.0139 *
## pangolin1$yearyear2015   -38.62      26.06  -1.482   0.1548
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 45.72 on 19 degrees of freedom
## Multiple R-squared:  0.2809, Adjusted R-squared:  0.2052
## F-statistic: 3.711 on 2 and 19 DF,  p-value: 0.04359
```

One-way ANOVA case study

- We found: the intercept is the mean in the first group (`year2013`) (baseline)
- Two others describe the **difference** between `year2014` and `year2015` to `year2013`.
- We haven't known the comparison between the second and the third group?

One-way ANOVA case study

- Use **multiple testing**.
- Conducting many tests can increase the probability of finding one of them to be significant
- Use an adjustment method, **the Bonferroni correction**

One-way ANOVA case study

```
pairwise.t.test(pangolin1$scales, pangolin1$year, p.adj="bonferroni")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  pangolin1$scales and pangolin1$year
##
##          year2013 year2014
## year2014 0.042    -
## year2015 0.464    1.000
##
## P value adjustment method: bonferroni
```

- We found there is difference between `year2013` and `year2014`
- No difference between `year2015` with `year2013` nor `year2014`.

One-way ANOVA case study

- We can relax the assumption of equal variances for all groups by using the `oneway.test`

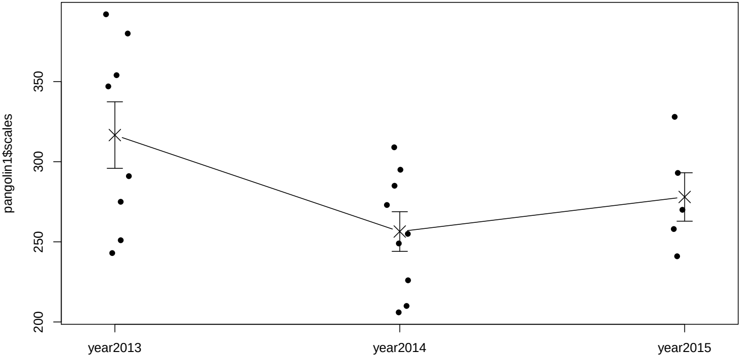
One-way ANOVA case study

```
oneway.test(pangolin1$scales~pangolin1$year)
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  pangolin1$scales and pangolin1$year
## F = 2.9704, num df = 2.000, denom df = 11.065, p-value = 0.09277
```

- We now see that the p-value has increased to a non-significant values
- Maybe, the group that seems to differ from the two others also has the largest variance.

One-way ANOVA case study



Plotting: which error bar to use?

- Should we draw using 1 SEM (done here) or using CIs ($\sim$ 2 SEM?)
- Or maybe SD instead of SEM?
- SEM = not useful for describing the distributions in the groups; only say how precise the mean is.
- SDs is difficult to see significant difference.
- Traditionally, people use 1 SEM (more dramatic)

Bartless test for checking variance

- **Bartlett's test**: to test whether the distribution of a variable has the same variance in all groups.

```
bartlett.test(pangolin1$scales~pangolin1$year)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  pangolin1$scales by pangolin1$year
## Bartlett's K-squared = 2.0951, df = 2, p-value = 0.3508
```

- No violation of equal variances assumptions in the three groups.

Two-way ANOVA

- When we have more than two factors to test on a dependent variable: two-way ANOVA.
- Single observation per cell of data: **residual variation** SSD_res = substraction of **variation between rows** (SSD_R) and **variation between columns** (SSD_C).

Two-way ANOVA case study

- A fake dataset: mean daily catch of Silky shark (*Carcharhinus falciformis*) in January 2017.
- We want to see the difference of catch by area (Meulaboh, Sigli and Biereun) (*factor 1*) depending on the whether the shark is female/male (*factor 2*. * The real dataset comes from here.

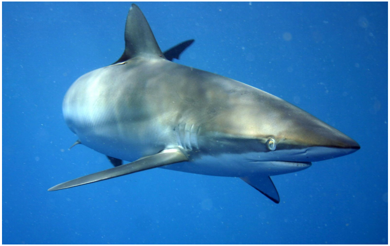


Figure 4: Silky shark

Two-way ANOVA case study

```
## [1] "sex"           "area"          "catch.per.day"
```

- We can then conduct the two-way ANOVA as:

```
calci.aov<-aov(catch.per.day~sex+area,data=calci)
summary(calci.aov)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## sex          1  0.067  0.0666   0.087  0.770
## area         2  0.278  0.1392   0.181  0.835
## Residuals   32 24.575  0.7680
```

- We found: both species sex and area are not statistically significant to the catch.

Two-way ANOVA case study

- To see if there is an interaction effect:

```
calci2.aov<-aov(catch.per.day~sex+area+sex:area,data=calci)
summary(calci2.aov)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## sex          1  0.067  0.0666   0.084  0.774
## area         2  0.278  0.1392   0.176  0.840
## sex:area     2  0.808  0.4038   0.510  0.606
## Residuals   30 23.767  0.7922
```

- We do not find any significant interaction.

Two-way ANOVA case study

We can compute some summary statistics (mean, sd, se):

```
model.tables(calci.aov, type="means", se=TRUE)
```

```
## Tables of means
## Grand mean
##
## 3.170833
##
##  sex
##     female  male
##      3.201  3.11
## rep 24.000 12.00
##
##  area
##     bireuen meulaboh  sigli
##       3.291    3.083  3.139
## rep  12.000   12.000 12.000
```

Two-way ANOVA case study

```
model.tables(calci2.aov, type="means", se=TRUE)
```

```
## Tables of means
## Grand mean
##
## 3.170833
##
##  sex
##     female  male
##      3.201  3.11
## rep 24.000 12.00
##
##  area
##     bireuen meulaboh  sigli
##       3.291    3.083  3.139
## rep  12.000   12.000 12.000
##
##  sex:area
##         area
## sex      bireuen meulaboh sigli
##   female 3.235   3.050    3.319
##   rep    8.000   8.000    8.000
##   male   3.403   3.148    2.780
##   rep    4.000   4.000    4.000
```

Two-way ANOVA case study

- We compute Tukey HSD (Honest Significant Differences) to perform multiple pairwise-comparison between the means of groups.
- We have seen that no difference between the groups, but let's give it a try:

```
TukeyHSD(calci.aov, which = "area")
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = catch.per.day ~ sex + area, data = calci)
##
## $area
##                         diff        lwr       upr     p adj
## meulaboh-bireuen -0.20833333 -1.0874842 0.6708175 0.8304934
## sigli-bireuen    -0.15166667 -1.0308175 0.7274842 0.9059786
## sigli-meulaboh    0.05666667 -0.8224842 0.9358175 0.9862700
```

Two-way ANOVA case study

```r
TukeyHSD(calci2.aov, which = "area")
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = catch.per.day ~ sex + area + sex:area, data = calci)
##
## $area
##                         diff        lwr       upr     p adj
## meulaboh-bireuen -0.20833333 -1.1041435 0.6874768 0.8352605
## sigli-bireuen    -0.15166667 -1.0474768 0.7441435 0.9087323
## sigli-meulaboh    0.05666667 -0.8391435 0.9524768 0.9866881
```

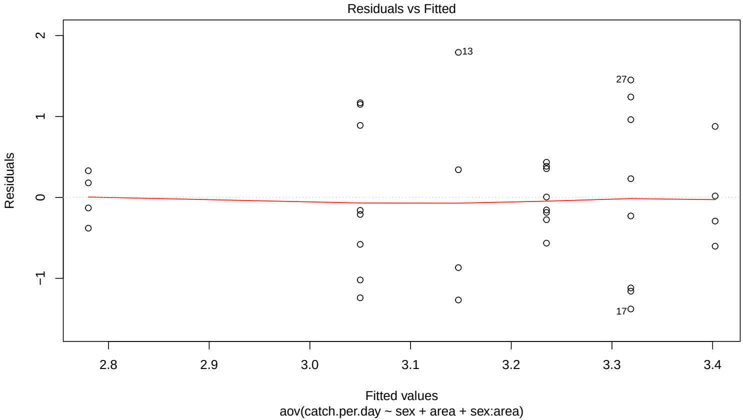
Checking assumptions of ANOVA

- Assumptions to check before running ANOVA (or any test assuming normal distribution):

1. Checking the homogeneity of variance assumptions

Checking assumptions of ANOVA

```
plot(calci2.aov, 1)
```



Points 13, 27 and 17 are detected as outliers, which can severely affect normality and homogeneity of variance. It can be useful to remove outliers to meet the test assumptions.

Checking assumptions of ANOVA

- Use Levene's test to check the homogeneity of variances.
- `leveneTest()` in car package will be used:

```
library(car)
leveneTest(catch.per.day~sex*area,data=calci)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value  Pr(>F)
## group  5  3.4628 0.01374 *
##       30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- p-value $< 0.05 =$ evidence to suggest group variances are different.
- We cannot assume homogeneity of variances.

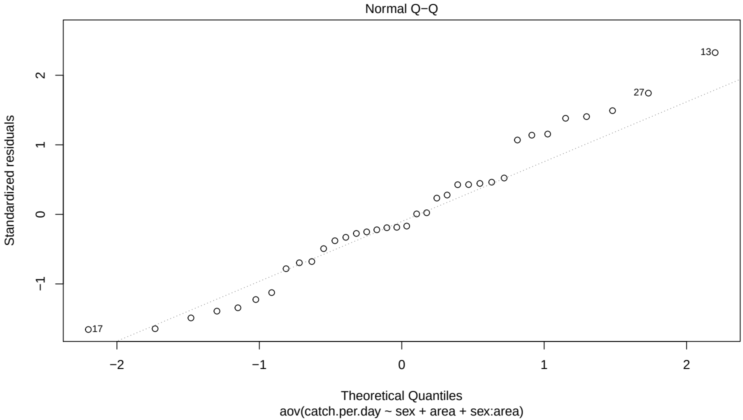
Checking assumptions of ANOVA

2. Check the normality assumption

- Use normal probability plot of residuals
- Should approximately follow a straight line.

Checking assumptions of ANOVA

```
plot(calci2.aov, 2)
```



- All points fall along the reference line: can assume normality.

Checking assumptions of ANOVA

- Run a **Shapiro-Wilk test** on the ANOVA residuals:

```
# Extract the residuals
aov_residuals <- residuals(object = calci2.aov)
# Run Shapiro-Wilk test
shapiro.test(x = aov_residuals )
```

```
##
##  Shapiro-Wilk normality test
##
## data:  aov_residuals
## W = 0.97011, p-value = 0.4284
```

- p-value is >0.05: do not find violation of normality.

Unbalanced design

- Our example above: deals with unbalanced design/not having equal numbers of subjects in each group.
- Three different ways to run an ANOVA: Type-I, Type-II and Type-III (recommended) sums of squares.

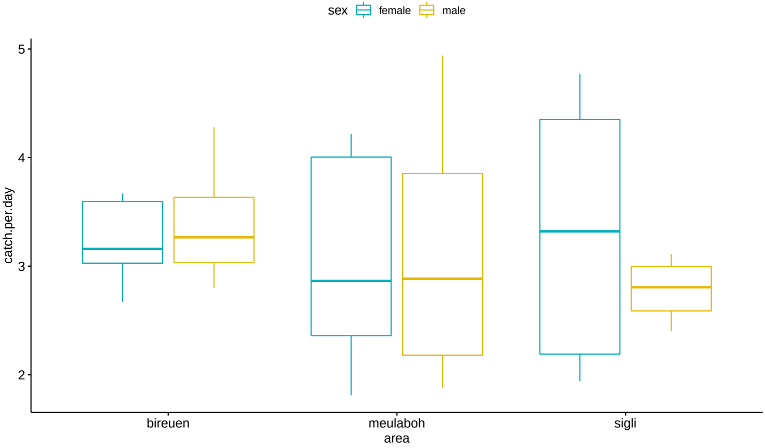
```
calci.unbal <- aov(catch.per.day~sex*area,data=calci)
Anova(calci.unbal, type = "III")
```

```
## Anova Table (Type III tests)
##
## Response: catch.per.day
##             Sum Sq Df  F value    Pr(>F)
## (Intercept) 83.722  1 105.6780 2.396e-11 ***
## sex          0.075  1   0.0944    0.7607
## area         0.303  2   0.1910    0.8272
## sex:area     0.808  2   0.5097    0.6058
## Residuals   23.767 30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- We find similar results, but already acknowledge the unbalanced design.

Plotting result

Overview of the data set, separated into the different sex and area by:

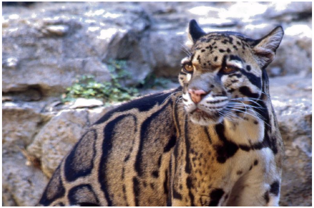


ANCOVA (analysis of covariance)

- Used to test **main and interaction** effects of **categorical variables** on a **continuous dependent variable**
- Controlling for other continuous variables: co-vary with the dependent.
- The control variables: “covariates.”
- No covariates: use ANOVA. Yes covariates: use ANCOVA.

ANCOVA case study

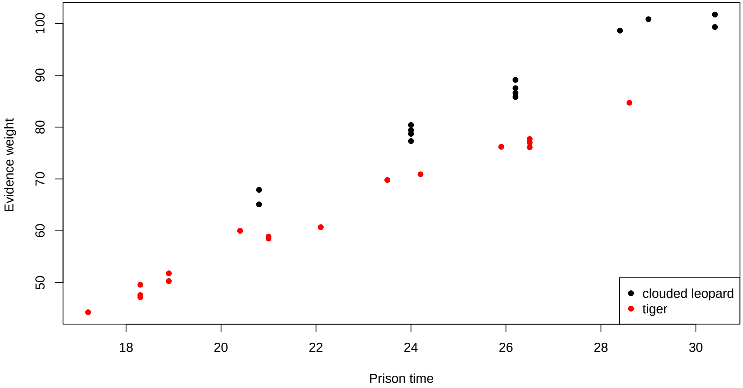
- Fake data: how long does a criminal got prison sentence? Variables: weight (kg skin) of confiscated evidence (by species: tiger or clouded leopard).



Clouded leopard & Indian Tiger.

ANCOVA case study

```
##           species evidence.weight prison.time
## 1 clouded leopard            20.8        67.9
## 2 clouded leopard            20.8        65.1
## 3 clouded leopard            24.0        77.3
```



ANCOVA case study

```
model.1 = lm (prison.time ~ evidence.weight + species + evidence.weight:species,
              data = confs)
Anova(model.1, type="II")
```

```
## Anova Table (Type II tests)
##
## Response: prison.time
##                          Sum Sq Df  F value    Pr(>F)
## evidence.weight          4376.1  1 1388.839 < 2.2e-16 ***
## species                   598.0  1  189.789 9.907e-14 ***
## evidence.weight:species     4.3  1    1.357    0.2542
## Residuals                  85.1 27
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Interaction = not significant; slope across groups = not different.

ANCOVA case study

- Re-run the model: is there difference among different species (without interaction):

```
model.2 = lm (prison.time ~ evidence.weight + species,
              data = confs)
Anova(model.2, type="II")
```

```
## Anova Table (Type II tests)
##
## Response: prison.time
##                  Sum Sq Df F value    Pr(>F)
## evidence.weight 4376.1  1  1371.4 < 2.2e-16 ***
## species          598.0  1   187.4 6.272e-14 ***
## Residuals         89.3 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Species is significant; intercepts among groups are different.

ANCOVA case study

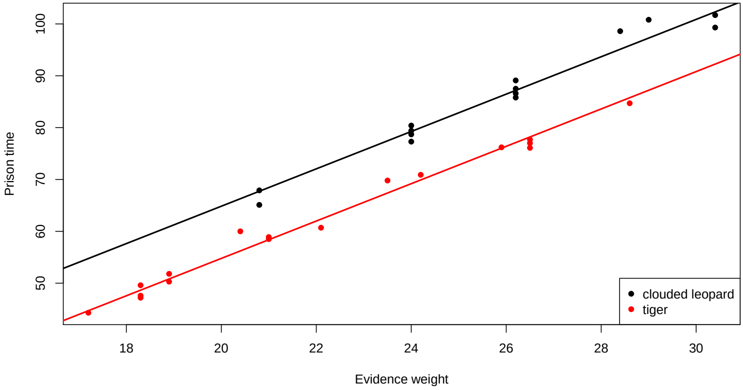
```
summary(model.2)
```

```
##
## Call:
## lm(formula = prison.time ~ evidence.weight + species, data = confs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0128 -1.1296 -0.3912  0.9650  3.7800
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -7.21091    2.55094  -2.827  0.00858 **
## evidence.weight   3.60275    0.09729  37.032  < 2e-16 ***
## speciestiger    -10.06529    0.73526 -13.689 6.27e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.786 on 28 degrees of freedom
## Multiple R-squared:  0.9896, Adjusted R-squared:  0.9888
## F-statistic:  1331 on 2 and 28 DF,  p-value: < 2.2e-16
```

- Slope estimate is the same.
- Intercept for species 1 (Clouded leopard) is `(Intercept)`.
- Intercept for species 2 (Tiger) is (intercept) + Speciestiger.
- Determined from the fact that Speciestiger is shown in coefficient table above.

ANCOVA case study

- Add the fitted lines:



Some notes on Type I, II, and III sums of squares:

- Type I = `anova` function.
- Type II and Type III = `Anova` in the `car` package
- Type I = "sequential." The factors tested in the order they are listed in the model.
- Type III = "partial." Every term in the model is tested after every other term in the model (interaction and other main effects)
- Type II are similar to Type III, but uses principle of marginality: test for main effects but not interaction.
- When data are balanced and the design is simple, types I, II, and III = same results.

Conclusions on ANOVA

- **One way ANOVA** $=$ one continuous variable (e.g., test score) compared by ≥ 3 levels of **one** factor variable (e.g., level of education)
- **Two-way ANOVA** $=$ one continous variable (e.g., test score) compared by * > 1 * factor variable (e.g., level of education and zodiac sign)
- **ANCOVA** $=$ *one continuous variable*, but it compares a response variable by *both* a *factor* and a *continuous* independent variable (e.g., comparing test score by 'level of education' and 'number of hours spent studying'). The continuous independent variable is 'covariate'.

One-way ANOVA

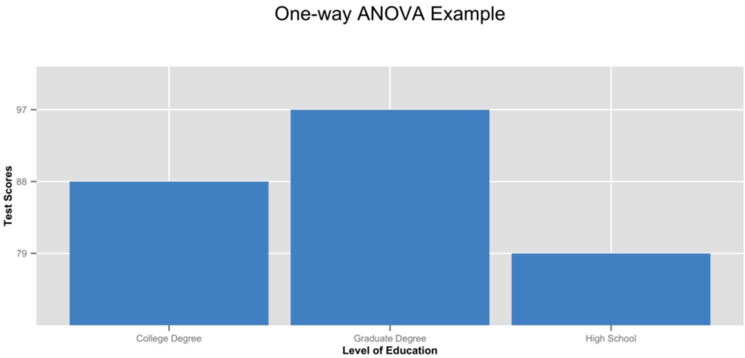


Figure 5: One-way ANOVA

Two way ANOVA

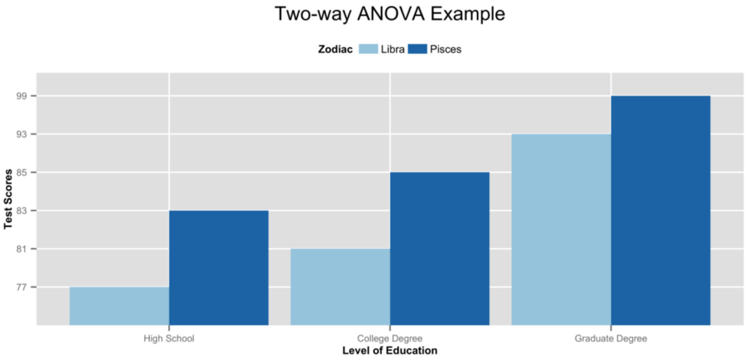


Figure 6: Two-way ANOVA

ANCOVA

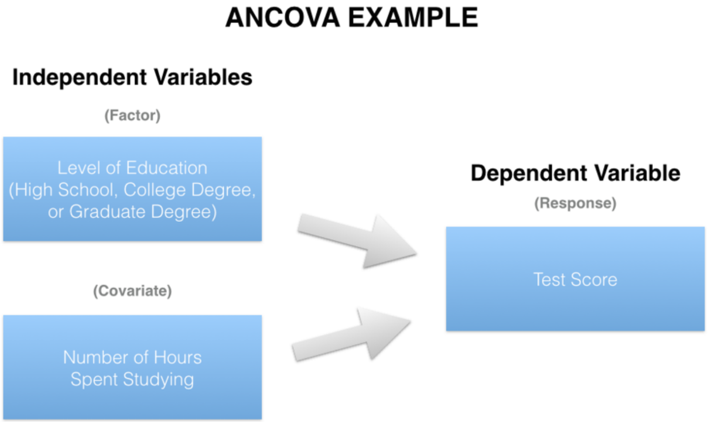


Figure 7: ANCOVA

Non-parametric test

Wilcoxon test : One-sample Wilcoxon

- Non-parametric version of **t-test**.
- Data are replaced with their ranks without regard to the grouping.
- **One-sample Wilcoxon** signed-rank test = done similarly to one-sample t-test.

```
#daily energy intake in kJ for 11 gorillas to be:
t.test(daily.intake,mu=7725) #parametric
```

```
##
##  One Sample t-test
##
## data:  daily.intake
## t = -2.8208, df = 10, p-value = 0.01814
## alternative hypothesis: true mean is not equal to 7725
## 95 percent confidence interval:
##  5986.348 7520.925
## sample estimates:
## mean of x
##  6753.636
```

Wilcoxon test : One-sample Wilcoxon

```r
wilcox.test(daily.intake,mu=7725) #non-parametric
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  daily.intake
## V = 8, p-value = 0.0293
## alternative hypothesis: true location is not equal to 7725
```

- Both tests show significant difference between the daily intake vs recommended intake (7725 kJ).

Two-sample Wilcoxon signed-rank test

- Similar to two-sample t-test

```
attach(energy)
#Differentiate the energy expend difference for obese gorilla vs lean gorilla.
t.test(expend~stature)#parametric
```

```
##
##  Welch Two Sample t-test
##
## data:  expend by stature
## t = -3.8555, df = 15.919, p-value = 0.001411
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.459167 -1.004081
## sample estimates:
##  mean in group lean mean in group obese
##            8.066154           10.297778
```

Two-sample Wilcoxon signed-rank test

```
wilcox.test(expend~stature)#non parametric
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  expend by stature
## W = 12, p-value = 0.002122
## alternative hypothesis: true location shift is not equal to 0
```

- Both tests show significant difference between obese vs lean gorillas energy expend.

Paired Wilcoxon test

- Similar to the paired t.test:

```
#Difference between the energy intake pre and post-release
#into the wild (after captured by hunters) in a group of gorillas.
attach(intake)
t.test(pre,post, paired=T) #parametric
```

```
##
##  Paired t-test
##
## data:  pre and post
## t = 11.941, df = 10, p-value = 3.059e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1074.072 1566.838
## sample estimates:
## mean of the differences
##                1320.455
```

- Both showed significant difference between the energy intake of pre and post release of gorillas into the wild.

Paired Wilcoxon test

```r
wilcox.test(pre,post, paired=T)#non parametric
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  pre and post
## V = 66, p-value = 0.00384
## alternative hypothesis: true location shift is not equal to 0
```

```r
detach(intake)
```

Kruskal Wallis test

- ‘One way ANOVA’ for non-parametric data.
- Compares differences between > 2 means.
- Similar to Wilcoxon-two sample test: data are replaced with their ranks, no regard to grouping.

```
oneway.test(pangolin1$scales~pangolin1$year) #parametric
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  pangolin1$scales and pangolin1$year
## F = 2.9704, num df = 2.000, denom df = 11.065, p-value = 0.09277
```

Kruskal Wallis test

```
kruskal.test(pangolin1$scales~pangolin1$year) #non-parametric
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  pangolin1$scales by pangolin1$year
## Kruskal-Wallis chi-squared = 4.1852, df = 2, p-value = 0.1234
```

- For both, we found no difference of pangolin scales confiscated evidence along the years.

Chi-square test of independence

- Used to analyze the frequency table (i.e. contingency table)
- Two categorical variables.
- Test significant association between two categorical variables.

Chi-square test of independence

A boring contingency table

- Here is an exceptionally boring contingency table, showing how a deck of cards is divided into suits and colors:

	Diamond	Heart	Club	Spade	Total
Red	13	13	0	0	26
Black	0	0	13	13	26
Total	13	13	13	13	52

- Note that the row totals and column totals both sum to the grand total of 52.

Figure 8: Contingency table example

Chi-square test of independence

- Are these songbirds associated with certain provinces? (ownership)

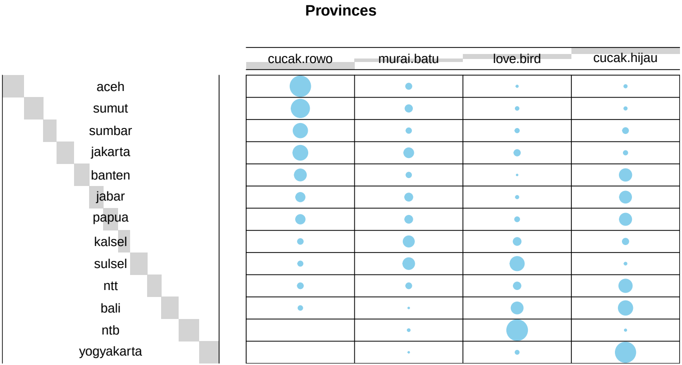
Cucak rowo, murai batu, lovebird and cucak hijau respectively.

Chi-square test of independence

```
songbird<-read.csv("fake_songbird.csv",row.names = 1)
```

- Can be visualized using **balloonplot()** (`gplots` package).
- Draws graphical matrix, each cell contains a dot whose size reflects the relative magnitude of the corresponding component.

Chi-square test of independence



Chi-square test of independence

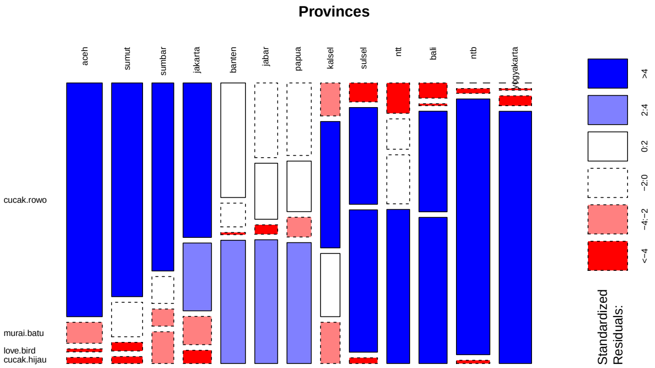
- We found: cucak rowo corresponding strongly to Aceh, Sumut, Sumbar and Jakarta.
- Need further tests to see whether this association is significant or not.

Chi-square test of independence

- Visualize using mosaic plot.
- Use `mosaicplot()`

Chi-square test of independence

```
mosaicplot(dt, shade = TRUE, las=2,
        main = "Provinces")
```



Chi-square test of independence

- Conclusions:
  - *Blue color* = observed value > expected value; if data were random.
  - *Red color* = specifies that observed value < expected value; if data were random
  - Cucak Rowo is mainly kept as a pet in Aceh, Sumut, Sumbar and Jakarta.

Chi-square test of independence

- Examines whether rows and columns of a contingency table are **associated**.
  - H_0: row and the column variables of the contingency table are independent.
  - H_1: row and column variables are dependent

Chi-square test of independence

- For each cell, expected value is calculated:

$$e = \frac{row.sum \times col.sum}{grand.total}$$

Chi-square test of independence

- Chi-square statistic is calculated as:

$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

Chi-square test of independence

- Chi-square statistic is compared to the critical value (obtained from statistical tables)
- $df = (r-1)(C-1)$ degrees of freedom and p $= 0.05$.
- Calculated Chi-square statistic $>$ critical value $=$ row and the column variables are not independent of each other (associated)

Chi-square test of independence case study

```
chisq<-chisq.test(songbird)
chisq
```

```
##
##  Pearson's Chi-squared test
##
## data:  songbird
## X-squared = 1944.5, df = 36, p-value < 2.2e-16
```

- Row and the column variables are statistically significantly associated (p-value $= 0$).

Chi-square test of independence case study

- Observed and the expected counts can be extracted:

```
chisq$observed #observed counts
```

```
##            cucak.rowo murai.batu love.bird cucak.hijau
## aceh              156         14         2           4
## sumut             124         20         5           4
## sumbar             77         11         7          13
## jakarta            82         36        15           7
## banten             53         11         1          57
## jabar              32         24         4          53
## papua              33         23         9          55
## kalsel             12         46        23          15
## sulsel             10         51        75           3
## ntt                13         13        21          66
## bali                8          1        53          77
## ntb                 0          3       160           2
## yogyakarta          0          1         6         153
```

Chi-square test of independence case study

```r
chisq$expected #expected counts
```

```
##            cucak.rowo murai.batu love.bird cucak.hijau
## aceh         60.55046   25.63303  38.44954    51.36697
## sumut        52.63761   22.28326  33.42489    44.65424
## sumbar       37.15596   15.72936  23.59404    31.52064
## jakarta      48.16514   20.38991  30.58486    40.86009
## banten       41.97248   17.76835  26.65252    35.60665
## jabar        38.87615   16.45757  24.68635    32.97993
## papua        41.28440   17.47706  26.21560    35.02294
## kalsel       33.02752   13.98165  20.97248    28.01835
## sulsel       47.82110   20.24427  30.36640    40.56823
## ntt          38.87615   16.45757  24.68635    32.97993
## bali         47.82110   20.24427  30.36640    40.56823
## ntb          56.76606   24.03096  36.04644    48.15654
## yogyakarta   55.04587   23.30275  34.95413    46.69725
```

Chi-square test of independence case study

- To know the most contributing cells to the total Chi-square score:

$$r = \frac{o - e}{\sqrt{e}}$$

Chi-square test of independence case study

- Previous formula returns Pearson residuals (r) for each cell (or standardized residuals).
- Cells with highest absolute standardized residuals contribute most to total Chi-square score.
- Pearson residuals can be easily extracted from the output of the function chisq.test():

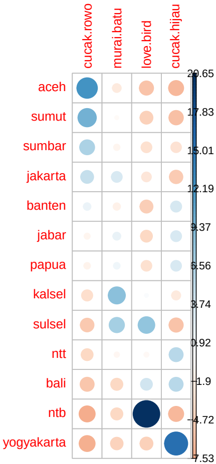
```
chisq$residuals
```

```
##            cucak.rowo murai.batu  love.bird cucak.hijau
## aceh        12.266344 -2.2976972 -5.8782288   -6.608968
## sumut        9.836062 -0.4836881 -4.9165873   -6.083794
## sumbar       6.536557 -1.1924678 -3.4162600   -3.298819
## jakarta      4.875263  3.4569876 -2.8180566   -5.297104
## banten       1.702142 -1.6056807 -4.9689076    3.585198
## jabar       -1.102817  1.8592103 -4.1634703    3.486107
## papua       -1.289342  1.3211002 -3.3623444    3.375634
## kalsel      -3.658896  8.5628768  0.4427320   -2.459431
## sulsel      -5.469207  6.8355754  8.0996318   -5.898312
## ntt         -4.150095 -0.8522912 -0.7419395    5.749805
## bali        -5.758422 -4.2771092  4.1073055    5.719883
## ntb         -7.534325 -4.2901609 20.6456123   -6.651286
## yogyakarta  -7.419291 -4.6201370 -4.8973513   15.556033
```

Chi-square test of independence case study

- Visualize Pearson residuals

```
library(corrplot)
corrplot(chisq$residuals, is.cor=FALSE)
```



Chi-square test of independence case study

- Sign of the standardized residuals = very important to interpret association
  - Positive residuals (blue): attraction (positive association)
  - Results: association between Cucak Rowo and Aceh, Sumut, Sumbar, Jakarta
  - Strong positive association between murai baru and Kalses
  - Negative residuals (red) = repulsion (negative association).
  - Cucak Rowo are negatively associated ($\sim$ "not associated") with the row Yogyakarta.
  - Repulsion between Love bird and with Aceh and Sumut.

Chi square goodness of fit test in R

- Used to compare the observed distribution to an expected distribution
- Compares multiple observed proportions to expected probabilities.

Chi square goodness of fit test in R case study

- Stingray species (*Dasyatis americana*) in Tj. Luar were found in January 2017: 81 juveniles, 50 adults and 27 old.

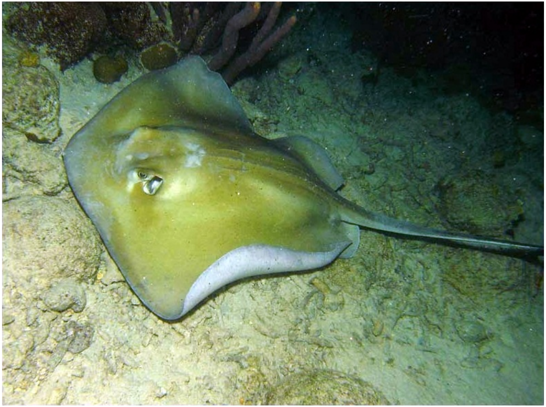


Figure 9: Stingray

Chi square goodness of fit test in R case study

- *Question 1:*
  - Are these age group equally common?
  - If these age groups were equally distributed, the expected proportion would be 1/3 for each age group.

Chi square goodness of fit test in R case study

- *Question 2:*
- Suppose that, in Indonesia ratio of juvenile, adults and old is 3:2:1 ($3+2+1 = 6$).
- Expected proportion is:
  - $3/6$ ($= 1/2$) for juvenile
  - $2/6$ ($= 1/3$) for adults
  - $1/6$ for old
- Is there significant difference between observed and expected proportions?

Chi square goodness of fit test in R case study

- H_0: No significant difference between observed and expected value.
- H_a: There is significant difference between observed and expected value.

Chi square goodness of fit test in R case study

- Q1: are the age groups equally common?

```
age<-c(81,50,27)
res<-chisq.test(age, p = c(1/3, 1/3,1/3)) #expected
```

- p-value is 8.80310^{-7} (p-value < 0.05).
- Age groups are significantly not commonly distributed.

Chi square goodness of fit test in R case study

- Access to the expected values:

```
res$expected
```

```
## [1] 52.66667 52.66667 52.66667
```

Chi square goodness of fit test in R case study

Q2: comparing observed to expected proportions

```
age <- c(81, 50, 27)
res <- chisq.test(age, p = c(1/2, 1/3, 1/6))
res
```

```
##
##  Chi-squared test for given probabilities
##
## data:  age
## X-squared = 0.20253, df = 2, p-value = 0.9037
```

- p-value = 0.9037
- Observed proportions are not significantly different from the expected proportions.

Correlation does not mean causation!

- Inability to deduce a cause-and-effect relationship between two variables solely on the basis of an **observed association** or **correlation** between them.
- Logical fallacy: how two events occurring together are considered to be a cause-and-effect relationship.

Correlation does not mean causation!

Children that watch a lot of TV are the most violent. Clearly, TV makes children more violent.
This could easily be the other way round; that is, violent children like watching more TV than less violent ones.

Correlation does not mean causation!

- Europeans in the Middle Ages believed that lice were beneficial to health -> rarely any lice on sick people -> people got sick because the lice left.
- Lice are extremely sensitive to body temperature
- Impression that lice left before the person got sick.

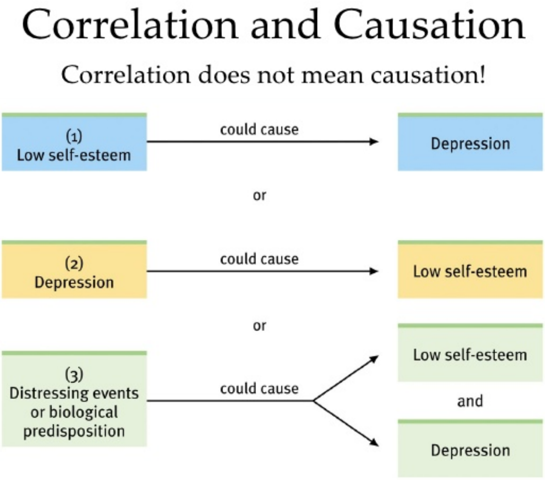
Correlation does not mean causation!

As ice cream sales increase, the rate of drowning deaths increases sharply. Therefore, ice cream consumption causes drowning.

Correlation does not mean causation!

Since the 1950s, both the atmospheric CO2 level and obesity levels have increased sharply. Hence, atmospheric CO2 causes obesity. Richer populations tend to eat more food and produce more CO2.

Correlation does not mean causation!



Practice!

- What test to use on this dataset?
  - "fake_deforestation.csv"
  - % soil organic carbon content (numeric) & deforestation rate (3 levels)
  - check the assumption, then decide which test!

Original data from see reference list.

Practice!

- What test to use on this dataset?
  - "fake.pangolin.csv"
  - Years of experience of the criminal, origin (province) of the criminal, weight of pangolin scale confiscated
  - check the assumption, then decide which test!
  - This is originally dataset `ToothGrowth` in R.

Practice!

- What test to use on this dataset?
  - “fake.manta.catch.csv”
  - Manta ray’s total body length (TL in cm) is measured a year before MPA, and a year after MPA. Is there any difference?
  - check the assumption, then decide which test!
  - Original dataset, check reference list.

Quiz 1

- A parametric test assumes:
  - no distribution of observed data
  - quantiles of observed data
  - standard error of observed data
  - normal distribution of observed data

Answer Quiz 1

- A parametric test assumes:
  - no distribution of observed data
  - quantiles of observed data
  - standard error of observed data
  - **normal distribution of observed data**

Quiz 2

- If you want to know whether a female lion runs faster than male lion, and you know that the data collected is normally distributed, you would use:
  - Kruskall Wallis test
  - t-test
  - ANCOVA test
  - chi-squared test

Answer Quiz 2

- If you want to know whether a female lion runs faster than male lion, and you know that the data collected is normally distributed, you would use:
  - Kruskall Wallis test
  - **t-test**
  - ANCOVA test
  - chi-squared test

Quiz 3

- If you want to know whether a grouper fish caught in Sigli, Tj Luar, and Kepulauan Seribu are different in body length, and you do not make any assumption of the data distribution, you would use:
  - Kruskall Wallis test
  - t-test
  - ANCOVA test
  - chi-squared test

Answer Quiz 3

- If you want to know whether a grouper fish caught in Sigli, Tj Luar, and Kepulauan Seribu are different in body length, and you do not make any assumption of the data distribution, you would use:
  - **Kruskall Wallis test**
  - t-test
  - ANCOVA test
  - chi-squared test

Quiz 4

- If you want to know whether a lovebird is sold due to its different sizes and or due to the area the shop is located, and you assume normal distribution of the observed data you would use:
  - Kruskall Wallis test
  - t-test
  - ANCOVA test
  - chi-squared test

Answer Quiz 4

- If you want to know whether a lovebird is sold due to its different sizes and or due to the area the shop is located, and you assume normal distribution of the observed data you would use:
  - Kruskall Wallis test
  - t-test
  - **ANCOVA test**
  - chi-squared test

Quiz 5

- If you want to know whether a survey on whether female or male eat more sharks depends on the education level of female/male, then we could use:
  - Kruskall Wallis test
  - t-test
  - ANCOVA test
  - chi-squared test

Answer Quiz 5

- If you want to know whether a survey on whether female or male eat more sharks depends on the education level of female/male, then we could use:
  - Kruskall Wallis test
  - t-test
  - ANCOVA test
  - **chi-squared test**

References

By KendallVarent - Own work, Public Domain,
https://commons.wikimedia.org/w/index.php?curid=10349551
https://en.wikipedia.org/wiki/Q%E2%80%93Q_plot
https://stats.idre.ucla.edu/spss/seminars/introduction-to-regression-with-spss/introreg-lesson2/
https://uc-r.github.io/assumptions_homogeneity
https://www.cscu.cornell.edu/news/statnews/stnews94.pdf
Zuur, et al. 2009. Especially for different data distributions and GLM tests.
Dalgaard. 2008. Especially on basic statistics and classical tests.
https://rmarkdown.rstudio.com/articles_intro.html for R markdown
https://www.dummies.com/education/math/statistics/what-a-p-value-tells-you-about-statistical-data/
https://effectsizefaq.com/category/p-values/
https://datascienceplus.com/standard-deviation-vs-standard-error/
https://serialmentor.com/dataviz/visualizing-uncertainty.html
https://stats.libretexts.org/Bookshelves/Biostatistics/Book%3A_Biological_Statistics_(McDonald)/3.0/3.3%3A_Standard_Error_of_the_Mean
https://www.leeds.ac.uk/educol/documents/00002182.htm
https://bookdown.org
http://www.statsmakemecry.com/smmctheblog/stats-soup-anova-ancova-manova-mancova
http://www.sthda.com/english/wiki/two-way-anova-test-in-r
https://www.iucn.org/sites/dev/files/styles/850x500_no_menu_article/public/deforestation_in_myanmar.jpg

https://www.sheffield.ac.uk/polopoly_fs/1.714570!/file/stcp-karadimitriou-KW.pdf (origin deforestation

data) http://www.sthda.com/english/wiki/paired-samples-t-test-in-r (origin manta data)

References

https://rcompanion.org/rcompanion/e_04.html
http://www.sthda.com/english/wiki/chi-square-goodness-of-fit-test-in-r
https://static1.squarespace.com/static/4f5694c424aca8d4f8e69194/t/53e5ff58e4b0cf25d06bb29a/1407582054892/one-way.ANOVA.png
https://static1.squarespace.com/static/4f5694c424aca8d4f8e69194/t/53e5fe86e4b0879d899b4195/1407581841371/two-way+anova.png
https://static1.squarespace.com/static/4f5694c424aca8d4f8e69194/t/53e8b030e4b0879d899d55ba/1407758388142/ancova.png
https://slideplayer.com/slide/12498591/74/images/4/A%2Bboring%2Bcontingency%2Btable.jpg
https://id.wikipedia.org/wiki/Berkas:Straw_Headed_Bulbul.jpg
https://id.wikipedia.org/wiki/Berkas:White-rumped_Shama_Copsychus_malabaricus_by_Dr._Raju_Kasambe_DSCN7003_(14).jpg
https://en.wikipedia.org/wiki/Lovebird#/media/File:Agapornis_roseicollis_-eating_grass_seeds-8.jpg
https://en.wikipedia.org/wiki/Greater_green_leafbird#/media/File:Chloropsis_sonnerati-20030531.jpg
https://en.wikipedia.org/wiki/Pangolin#/media/File:Pangolin_borneo.jpg
https://en.wikipedia.org/wiki/Silky_shark#/media/File:Carcharhinus_falciformis_off_Cuba.jpg
https://www.slideshare.net/AjayMalpani/parametric-and-non-parametric-test
https://en.wikipedia.org/wiki/Whiptail_stingray#/media/File:Dasyatis_americana_bonaire.jpg
http://www.statsmakemecry.com/smmctheblog/stats-soup-anova-ancova-manova-mancova
https://sites.google.com/site/kaylauthepsychologyobjectives/grade-level/objective-5
https://en.wikipedia.org/wiki/Correlation_does_not_imply_causation#Examples_of_illogically_inferring_causation_from_correlation
https://www.iucn.org/resources/issues-briefs/deforestation-and-forest-degradation

kruskal two way anova t-test