

Statistical Methods for Plant Biology

PBIO 3150/5150

Anirudh V. S. Ruhil

April 5, 2016

The Voinovich School of Leadership and Public Affairs

Table of Contents

- 1 The Correlation Coefficient
- 2 Testing the Null Hypothesis of $\rho = 0$
- 3 Spearman's Rank Correlation

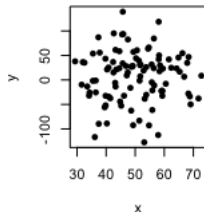
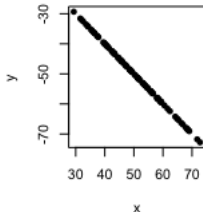
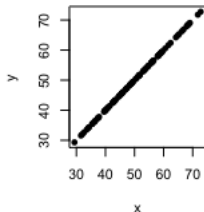
The Correlation Coefficient

The Correlation Coefficient r

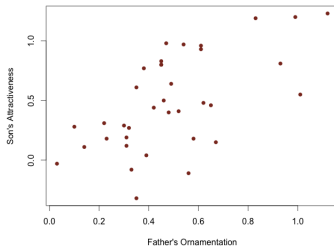
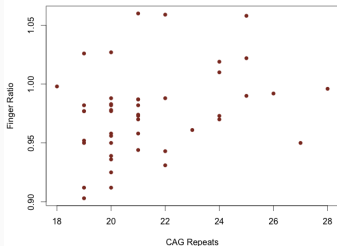
The correlation coefficient (r) estimates the association between two continuous (aka numerical) variables x and y

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}}$$

- $-1 \leq r \leq +1$
- $r = +1$ indicates a **perfect positive linear relationship**
- $r = -1$ indicates a **perfect negative linear relationship**
- $r \approx 0$ indicates a **absence of a linear relationship**



Some Examples



```
> cor(FingerRatio, use="complete.obs", method="pearson")
      CAGrepeats  finger.ratio
CAGrepeats  1.000000  0.308189
finger.ratio 0.308189  1.000000
>
> cor(Guppies, use="complete.obs", method="pearson")
      father.ornament  son.attract
father.ornament  1.0000000  0.6141043
son.attract     0.6141043  1.0000000
```

Testing the Null Hypothesis of $\rho = 0$

Testing r

Given that r is based on a sample it is estimating the true correlation between x and y in the population ... denoted by ρ

One then needs to conduct a statistical test that will tell us whether in the population $\rho = 0$ or $\rho \neq 0$ with $H_0: \rho = 0$; $H_A: \rho \neq 0$

The test statistic is: $t = \frac{r}{SE_r}$; where $SE_r = \sqrt{\frac{1-r^2}{n-2}}$

Reject H_0 if P -value of the calculated t is $\leq \alpha$; Do not reject H_0 otherwise

We can also calculate **asymptotic approximate confidence intervals** for ρ :

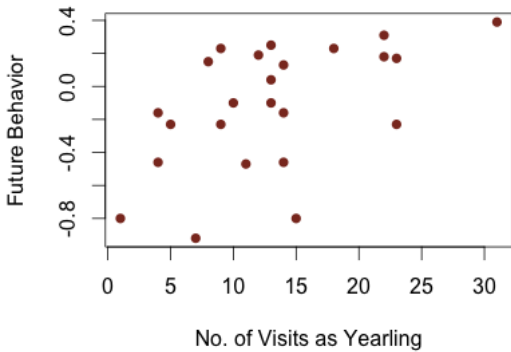
$z - 1.96\sigma_z < \zeta < z + 1.96\sigma_z$ where $z = 0.5 \ln\left(\frac{1+r}{1-r}\right)$; $\sigma_z = \sqrt{\frac{1}{n-3}}$; and ζ (zeta) is the population analogue of the z used to calculate confidence intervals

Because the z involves the natural logarithm we back-transform by taking the antilog of the lower and upper bounds of the confidence interval

Maltreatment and Youth Experience

Adults who mistreat children were often mistreated themselves when they were young. Is there a similar association in nonhuman animals?

Researchers investigated this possibility in the Nazca booby (*Sula granti*), a colonial nesting seabird of the Galapagos islands. Unattended chicks in nests frequently received visits from unrelated adults, who behaved mainly aggressively toward them. The researchers counted the number of such visits to nests of 24 booby chicks. These chicks were given unique numbered rings on their legs, which allowed the researchers to observe their behavior years later when they had become adults.



Hypothesis Testing & Confidence Intervals

```
> with(birds, cor.test(nVisitsNestling, futureBehavior))
      Pearson's product-moment correlation
data:  nVisitsNestling and futureBehavior
t = 2.9603, df = 22, p-value = 0.007229
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1660840 0.7710999
sample estimates:
      cor
0.5337225

> with(Guppies, cor.test(son.attract, father.ornament))
      Pearson's product-moment correlation
data:  Guppies$son.attract and Guppies$father.ornament
t = 4.5371, df = 34, p-value = 6.784e-05
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3577455 0.7843860
sample estimates:
      cor
0.6141043
```

Notice that you get the usual test results as well as the confidence intervals

In both there is a statistically significant positive correlation

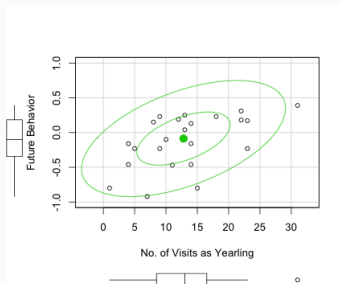
However, note how wide the confidence intervals are for each case

Assumptions and their Violations

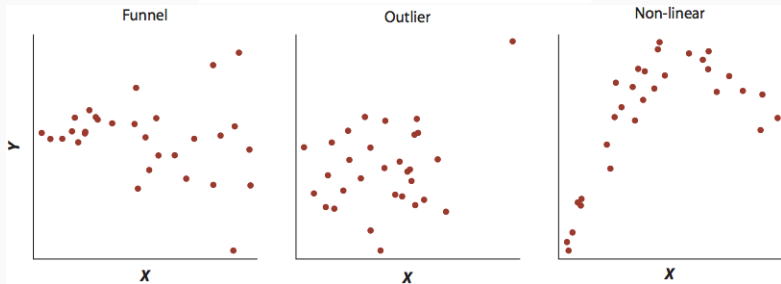
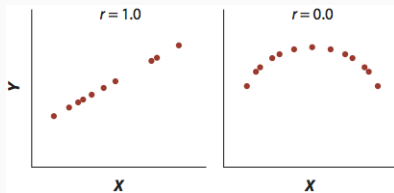
The Correlation Coefficient assumes **bivariate normality**

- x and y are jointly normally distributed
- x and y are linearly related
- The cloud of points has a circular or elliptical shape

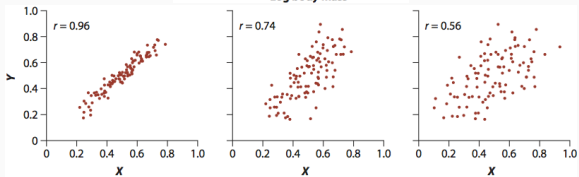
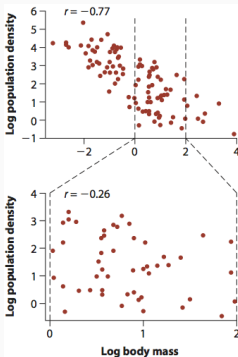
If these are violated we can try the usual transformations but if these fail we can then rely on a nonparametric approach. **Outliers? Bivariate normality is violated.**



Stylized Examples of Violations



Beware Attenuation and Measurement Error



Spearman's Rank Correlation

Spearman's Rank Correlation

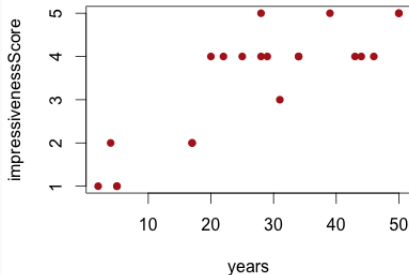
Measures strength and association between the **rank**s of two variables assumed to be (i) randomly sampled, and (ii) with linearly related ranks

- 1 Rank the scores of each variable separately, from low to high
- 2 Average the ranks in the presence of ties
- 3 Calculate $r_s = \frac{\sum (R - \bar{R})(S - \bar{S})}{\sqrt{\sum (R - \bar{R})^2} \sqrt{\sum (S - \bar{S})^2}}$
- 4 $H_0: \rho_s = 0; H_A: \rho_s \neq 0$
- 5 Set α
- 6 Reject H_0 if $P - \text{value} \leq \alpha$; Do not reject H_0 otherwise

The Indian Rope Trick

How reliable are witness accounts of “miracles”? One means of testing this is by comparing different accounts of extraordinary magic tricks. Of the many illusions performed by magicians, none is more renowned than the Indian rope trick. In brief, a magician tosses the end of a rope into the air and the rope forms a rigid pole. A boy climbs up the rope and disappears at the top. The magician scolds the boy and asks him to return but with no response, and so climbs the rope himself, with a knife in hand, and does not return. The boy’s body falls in pieces from the sky into a basket on the ground. The magician then drops back to the ground and retrieves the boy from the basket, revealing him to be unharmed and in one piece.

Researchers tracked down the 21 first-hand accounts and scored each narrative according to how impressive it was, on a scale of 1 to 5. The researchers also recorded the number of years that had lapsed between the date that the trick was witnessed and the date the memory of it was written down. Is there any association between the impressiveness of eyewitness accounts and the time lapsed since the account was penned?



```
> cor.test(RopeTrick$impressiveness, RopeTrick$years, method="spearman")
  Spearman's rank correlation rho
data: RopeTrick$impressiveness and RopeTrick$years
S = 332.1221, p-value = 2.571e-05
alternative hypothesis: true rho is not equal to 0
sample estimates: rho = 0.7843363
Warning message: In cor.test.default(RopeTrick$impressiveness, RopeTrick$years,
method = "spearman"): Cannot compute exact p-value with ties

> spearman_test(impressivenessScore ~ years, data = rope)
Z = 3.5077, p-value = 0.0004521
```