Statistical Methods for Plant Biology

PBIO 3150/5150

Anirudh V. S. Ruhil

March 15, 2016

The Voinovich School of Leadership and Public Affairs

Table of Contents

Graphing Departures from Normality

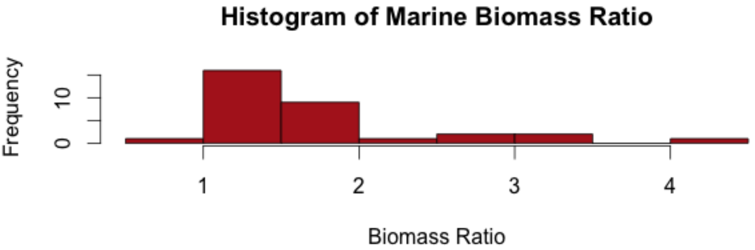
Marine Reserves

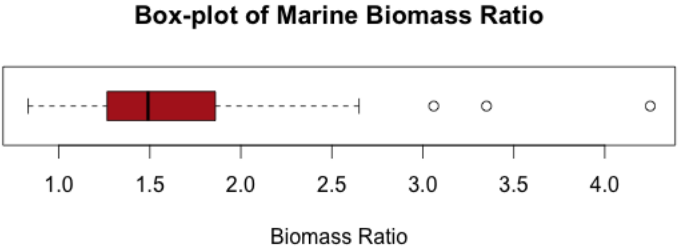
Marine reserves are becoming increasingly popular for biological conservation and the protection of fisheries. But are reserves effective in preserving marine wildlife?

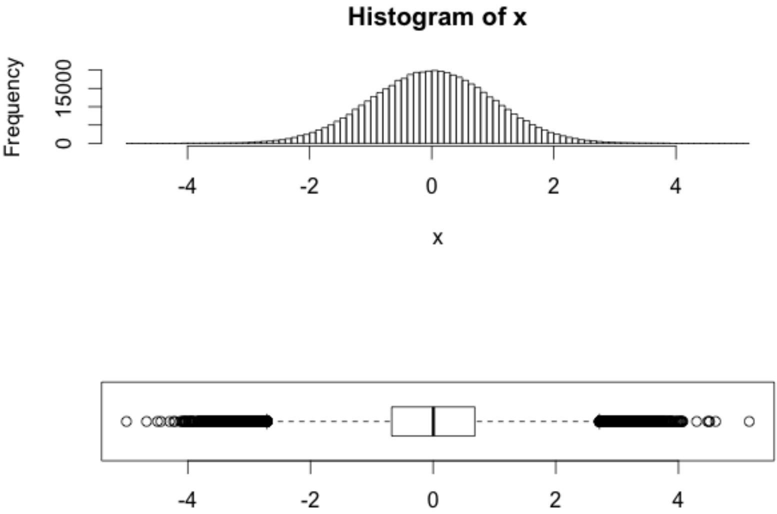
Halpern (2003) matched each of 32 marine reserves to a control location, which was either the site of the reserve before it became protected or a similar unprotected site nearby. One index of protection evaluated by the site was the “biomass ratio,” which is the total mass of all marine plants and animals per unit area of reserve divided by the same quantity in the unprotected control. This biomass ratio would equal one if protection had no effect. The biomass ratio would be > 1 if the protection were beneficial, and it would be < 1 if protection reduced biomass. Are marine reserves effective?

H_0 : The mean biomass ratio is unaffected by reserve protection $\mu = 1)$
H_A : The mean biomass ratio is affected by reserve protection $\mu \neq 1)$

We could use a a $t-test$ but only so long we can safely assume that the biomass ratios are drawn from a normal population. How can we test this assumption and what do we do if it is violated?

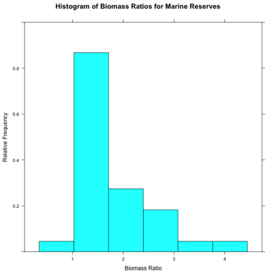

Histogram of Marine Biomass Ratio
Frequency
0
10
1
2
3
4
Biomass Ratio

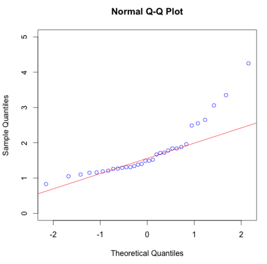

Box-plot of Marine Biomass Ratio
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Biomass Ratio


Histogram of x
Frequency
0
15000
-4
-2
0
2
4
x
-4
-2
0
2
4

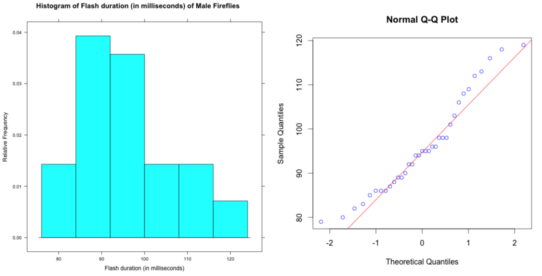
Graphical Methods

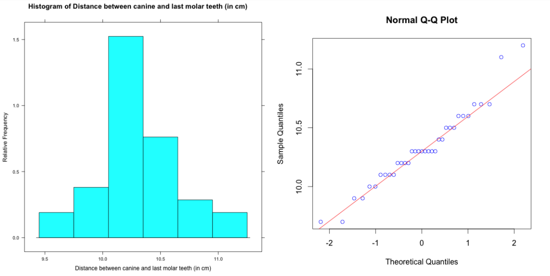
- Visually exploring your numerical measures via histograms or boxplots is a good way to see if distribution is Normal (or not)
- A very useful graphical device is the quantile-quantile plot … compares the observed distribution to what would be expected under a standard normal distribution (the $z - score$)
- Essentially the Quantile plot flips the data into a $z - score$ and then plots where each observation falls relative to what we know should hold if it were Normally distributed.
- If distribution is Normal all observed data should fall on a straight line; deviations from the straight line hint at violations of the Normality assumption





Some More Examples





Formal Test of Normality

A Formal Test: Shapiro-Wilk

H_0 : Sample comes from a $\sim N(.)$ population; H_A : Sample does not come from a $\sim N(.)$ population; Set $\alpha = 0.05$ or 0.01

Caution: You do not want to reject H_0 otherwise you may have to transform the data

- Shapiro-Wilk test statistic: $W = \dfrac{\left(\sum\limits^{n} a_i Y_{(i)}\right)^2}{\sum\limits^{n} (Y_i - \bar{Y})^2}; 0 \leq W \leq 1.$

  $W \to 1$: Observed distribution is as expected if population were Normal
- Skewness: $\sum\limits_{i=1}^{n} \dfrac{(Y_i - \bar{Y})^3}{(n-1)\, s^3}$

  $s = 0$: Normal distribution; $s < 0$: skewed left; $s > 0$: skewed right
- Kurtosis: $\sum\limits_{i=1}^{n} \dfrac{(Y_i - \bar{Y})^4}{(n-1)\, s^4} - 3$

  $k = 0$: Standard Normal distribution; $K > 0$: “Peaked”; $k < 0$: “Flat”

Some Examples

For MarineReserve *biomass.ratio*

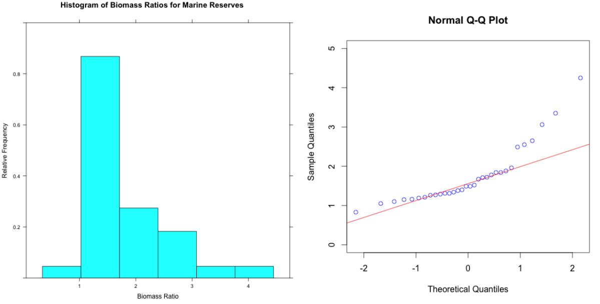
S-W: $W = 0.8175$,

$P - value =$

$8.851e - 05$

Skewness: $s = 1.61$

Kurtosis: $k = 2.29$



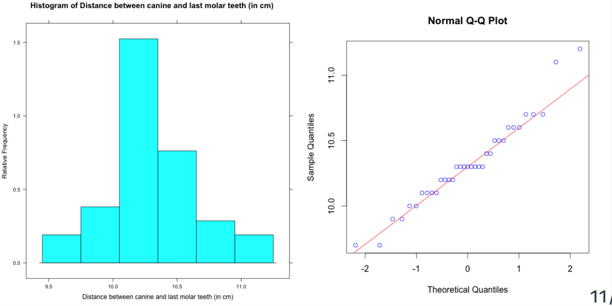
For WolfTeeth *length*

S-W: $W = 0.959$,

$P - value = 0.2133$

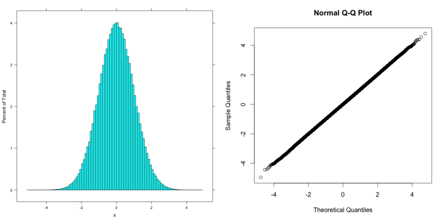
Skewness: $s = 0.50$

Kurtosis: $k = 0.39$



Caution!!

- H_0 : Sample comes from a Normally distributed population … Rejecting this does not necessarily mean that the sample does not come from a Normal distribution
- In small samples even large departures from Normality may be missed In large samples even small departures from Normality may be overstated
- Visual examination of plots may not be supported by formal tests



$A = 1.3584$, $P - value = 0.001617$ … hello??

- So what should you do? … we now turn to our options

1. Have faith in the Central Limit Theorem
   So long as I have a sample size of 30 or more I can assume that the sampling distribution of the sample mean follows the normal distribution (even if the original measurement does not come from a normal distribution)
2. Test statistics for means & confidence intervals of means will be unaffected so long as you have a "large enough" sample to work with
   - Most tests are robust to some violations of the normality assumption so long as the sample size is not too small and the skew isn't very extreme
   - How small? If there is similar skewness, even when comparing two groups, having 30 in each group will work
   - If I have one group with severe skewness (one left-skewed the other right-skewed) then I need samples of a few hundred units each before the test can be trusted
3. If all else fails (or you are a conservative analyst), try to transform the data and rerun all tests … We will try this tactic next
4. If this doesn't solve the problem (or you are not a fan of massaging your data) then switch to tests that do not require the Normality assumption … We will also see these tests in action

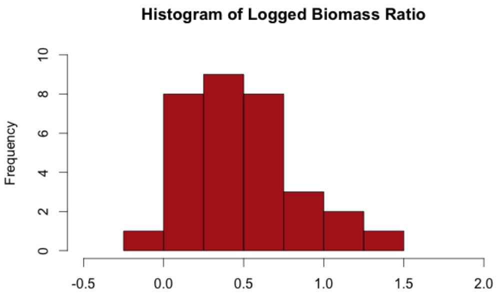
Data Transformations

Transformations

- If the original metric violates assumptions then try various transformations and test if the transformation works
  1. Natural logarithm used with substantial skewness:
     (a) very positively skewed and $Y > 0$ then use $Y^{'} = ln(Y)$
     (b) very positively skewed and $Y \geq 0$ then use $ln(Y + c)$; where c is a constant chosen such that the smallest score $= 1$
     (c) very negatively skewed then use $ln(K - Y)$; where $K = Y_{max} + 1$ so that the smallest score $= 1$
     (d) Often used with ratios or products of variables
  2. Square Root used with moderate skewness:
     (a) moderately positively skewed and $Y > 0$ then use $\sqrt{Y}$
     (b) moderately positively skewed and $Y \geq 0$ then use $\sqrt{Y + 0.5}$
     (c) moderately negatively skewed use $\sqrt{K - Y}$; where $K = Y_{max} + 1$ so that the smallest score $= 1$
     (d) Often used with count data
- Can also try the square: Y^2, the cube: Y^3, the antilog: e^Y, or the reciprocal: $\frac{1}{Y}$
- If data are proportions the arcsine: $arcsin\left(\sqrt{p}\right)$ is suggested but I would not do this readily ... Proportions are essentially coming from your Binomial or Poisson distributions and in these cases we would rarely run a $t - test$
- Moderate skew: $-1.0 \leq s \leq -0.5$ or $+0.5 \leq s \leq +1.0$
- Heavy skew: $s < -1.0$ or $s > +1.0$

Transforming Biomass Ratios

Since the data are all > 0 and right-skewed we can start with the natural logarithm of biomass ratios. The result?



Now we can run the $t - test$ with this logged version of biomass ratios but with the H_0 value being specified as $(\mu' = 0)$ and H_A as $(\mu' \neq 0)$ … because $ln[1] = 0$

H_0: Mean log biomass ratio is 0 $(\mu^{'} = 0)$
H_A: Mean log biomass ratio is not equal to 0 $(\mu^{'} = 0)$
$\alpha = 0.05$

```
> t.test(marine$ln.biomassRatio, mu=0, alternative="two.sided", conf.level=0.95)

       One Sample t-test
data:  marine$ln.biomassRatio
t = 7.3968, df = 31, p-value = 2.494e-08
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.3470180 0.6112365
sample estimates:
mean of x
0.4791272
```

Given the low $p - value$ we can reject H_0

Note: The confidence intervals are given in natural log units so if we take the antilog we can get the untransformed interval

```
> exp(c(0.3470180, 0.6112365))
[1] 1.414842 1.842708
```

Some Non-Parametric Tests

Non-Parametric Tests

- Non-parametric tests are used (a) when transformations do not work, or (b) the data represent ordinal categories (or are ranked data)
- Called non-parametric because unlike, say, the $t-test$ which requires some distributional assumption to be true (i.e., Normality) and involves parameters (i.e., the mean and the variance), these alternatives make no such assumptions or need no such parameters

  They are more likely to lead to a Type II error so if the assumptions of parametric tests are met use parametric tests
- You have already used a few related forms of a non-parametric test: The χ^2 Test, Fisher's Test
- Here are a few non-parametric analogues to the $t-tests$:

  1. Sign Test: Alternative to the One-Sample $t-test$ or to the Paired $t-test$
  2. Wilcoxon Signed-Rank Test: Alternative to the Paired $t-test$
  3. Mann-Whitney $U-test$: Alternative to the Two-Sample $t-test$ with equal variances
  4. Kolmogorov-Smirnov test: Alternative to the Two-Sample $t-test$ with unequal variances
  5. Permutation test: Alternative to the Paired and Two-sample t-test

The Sign Test

- Assumption: Independent samples from a continuous distribution
- Tests whether the Median equals a hypothesized value (H_0 value)
- Scores above H_0 value are marked +; scores below are marked -
- Scores $=$ to the Median are dropped
- If H_0 is correct, 50% of the scores should be "+" and 50% should be "-"
  ... essentially a Binomial test where $p_0 = 0.50$
- Hypotheses:
  H_0: Distribution is symmetric around $p = 0.50$ ... One-Sample test
  H_A: Distribution is not symmetric around $p \neq 0.50$

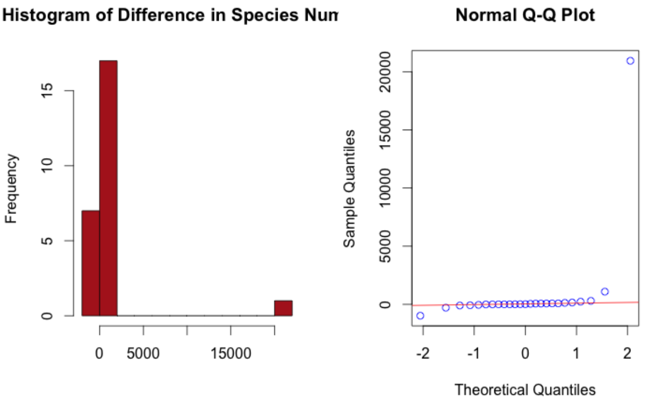
  H_0: Distribution of the two measurements is the same ... Paired test
  H_A: Distribution of the two measurements is not the same
- Has little power because it is simplistic. Works best with large samples

Sexual Conflict and the Origin of the Species

The process by which a single species splits into two species is still not well understood. One proposal involves “sexual conflict” – a genetic arms race between males and females that arises from their different reproductive roles.For example, in a number of insect species, male seminal fluid contains chemicals that reduce the tendency of the female to mate again with other males. However, these chemicals also reduce female survival, so females have developed mechanisms to counter these chemicals. Sexual conflict can cause rapid genetic divergence between isolated populations of the same species, leading to the formation of new species. Sexual conflict is more pronounced in species in which females mate more than once, leading to the prediction that they should form new species at a more rapid rate.

To investigate this, Arnqvist et al. (2000) identified 25 insect taxa (groups) in which females mate multiple times, and they paired each of these groups to a closely related insect group in which females only mate once. Which type of insects tend to have more species?

These are treated as paired data because the two sets of groups are closely related



```
> shapiro.test(taxa$difference)
        Shapiro-Wilk normality test
data:  taxa$difference
W = 0.2547, p-value = 2.911e-10
```

No. of Species					No. of Species				
polyandrous	monandrous	d_i	ID	Sign	polyandrous	monandrous	d_i	ID	Sign
53	10	43	A	$+$	37	115	-78	M	$-$
73	120	-47	B	$-$	100	30	70	N	$+$
228	74	154	C	$+$	21000	60	20940	O	$+$
353	289	64	D	$+$	37	40	-3	P	$-$
157	30	127	E	$+$	7	5	2	Q	$+$
300	4	296	F	$+$	15	7	8	R	$+$
34	18	16	G	$+$	18	6	12	S	$+$
3400	3500	-100	H	$-$	240	13	227	T	$+$
20	1000	-980	I	$-$	15	14	1	U	$+$
196	486	-290	J	$-$	77	16	61	V	$+$
1750	660	1090	K	$+$	15	14	1	W	$+$
55	63	-8	L	$-$	85	6	79	X	$+$
					86	8	78	Y	$+$

Note: There are 7 $-$ and 18 $+$ observations with $n = 25$

So 7 out of 25 independent trials generated a $-$

We would expect, under the assumption of no difference, 12 $-$ and 12 $+$

What, then, is the probability of observing 7 “successes” in 25 independent trials?

H_0: Median difference in number of species is $= 0$

H_A: Median difference in number of species is $\neq 0$

Set $\alpha = 0.05$

```
binom.test(7, 25, p=0.5)
       Exact binomial test
data:  x and n
number of successes = 7, number of trials = 25, p-value = 0.04329
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval: 0.1207167 0.4938768
sample estimates: probability of success = 0.28
```

Reject H_0; the Median difference is not $= 0$. The data suggest that groups of insects whose females mate multiple times have more species than groups whose females mate singly, consistent with the sexual-conflict hypothesis

The Wilcoxon Signed-Rank Test

- Assumes the distribution is symmetric around the Median ... $=$ no skew! so very restrictive
- Steps:
  1. Calculate $Y_i - \mu_0$ for all $i = 1, 2, \ldots, n$
  2. Rank in ascending order the absolute differences $|Y_i - \mu_0|$ for all $i = 1, 2, \ldots, n$
  3. Assign + or - to each rank
  4. Let the sum of + and - ranks be W^+ and W^-, respectively
  5. Let $W = min\left(W^+, W^-\right)$ and W^*_α be critical W
  6. Reject H_0: The median difference between the two samples is $\mu_0 = 0$ if ...
     - (a) H_A: is $\mu \neq \mu_0$ and $W \leq W^*_\alpha$
     - (b) H_A: is $\mu > \mu_0$ and $W^- \leq W^*_\alpha$
     - (c) H_A: is $\mu < \mu_0$ and $W^+ \leq W^*_\alpha$

polyandrous	monandrous	d_i	rank	sign	pair	polyandrous	monandrous	d_i	rank	sign	pair
15	14	1	1.5	+	U	37	115	-78	14.5	-	M
15	14	1	1.5	+	W	86	8	78	14.5	+	Y
7	5	2	3	+	Q	85	6	79	16	+	X
37	40	-3	4	-	P	3400	3500	-100	17	-	H
55	63	-8	5.5	-	L	157	30	127	18	+	E
15	7	8	5.5	+	R	228	74	154	19	+	C
18	6	12	7	+	S	240	13	227	20	+	T
34	18	16	8	+	G	196	486	-290	21	-	J
53	10	43	9	+	A	300	4	296	22	+	F
73	120	-47	10	-	B	20	1000	-980	23	-	I
77	16	61	11	+	V	1750	660	1090	24	+	K
353	289	64	12	+	D	21000	60	20940	25	+	O
100	30	70	13	+	N						

$W^+ = 1.5+1.5+3+5.5+7+8+9+11+12+13+14.5+16+18+19+20+22+24+25=230$

$W^- = 4+5.5+10+14.5+17+21+23=95$

```
wilcox.test(SexualSelection$polyandrous.species, SexualSelection$monandrous.species, paired=TRUE)
       Wilcoxon signed rank test with continuity correction
data:  SexualSelection$polyandrous.species and SexualSelection$monandrous.species
V = 230, p-value = 0.07139 alternative hypothesis: true location shift is not equal to 0
Warning message: In wilcox.test.default(SexualSelection$polyandrous.species,
      SexualSelection$monandrous.species, : cannot compute exact p-value with ties
```

Do not reject H_0; the distributions appear to be the same

The Mann-Whitney U-Test

- The assumptions of the Mann-Whitney U test are:
    1. The variable of interest is continuous (not discrete). The measurement scale is at least ordinal
    2. The probability distributions of the two populations are identical, except for location (i.e., the “center”)
    3. The two samples are independent
    4. Both are simple random samples from their respective populations
- H_0: The samples come from populations with similar probability distributions
- Test Process and Statistic ...
    1. Combine both samples and rank, in ascending order, all values
    2. If there are ties, rank accordingly
    3. Sum the ranks of the smaller group (R_1)
    4. $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1; U_2 = n_1 n_2 - U_1$
    5. Choose the larger of U_1 or U_2 as the test statistic
    6. Reject H_0 if $P-value \leq \alpha$

Sexual Cannibalism in Sagebrush Crickets

The sage cricket, *Cyphoderris strepitans*, has an unusual form of mating. During mating, the male offers his fleshy hind wings to the female to eat. The wounds are not fatal but a male with already nibbled wings is less likely to be chosen by females he meets later. Females get some nutrition from feeding on the wings, which raises the question, "Are females more likely to mate if they are hungry?"

Johnson et al. (1999) answered this question by randomly dividing 24 females into two groups: one group of 11 females were starved for at least two days and another group of 13 females was fed during the same period. Finally, each female was put separately into a cage with a single (new) male, and the waiting time to mating was recorded.

H_0: Time to mating is the same for female crickets that were starved as for those who were fed
H_A: Time to mating is not the same for female crickets that were starved as for those who were fed

treatment	time.to.mating	rank
fed	1.50	1
fed	1.70	2
starved	1.90	3
starved	2.10	4
fed	2.40	5
fed	3.60	6
starved	3.80	7
fed	5.70	8
starved	9.00	9
starved	9.60	10
starved	13.00	11
starved	14.70	12
starved	17.90	13
starved	21.70	14
fed	22.60	15
fed	22.80	16
starved	29.00	17
fed	39.00	18
fed	54.40	19
fed	72.10	20
starved	72.30	21
fed	73.60	22
fed	79.50	23
fed	88.90	24

The *starved* group is smaller so we'll work with their ranks

$R_1 = 3+4+7+9+10+11+12+13+14+17+21 = 121$

$U_1 = n_1 n_2 + \dfrac{n_1(n_1+1)}{2} - R_1 = 11(13)\dfrac{11(11+1)}{2} - 121$

$= 143 + \dfrac{132}{2} - 121 = 143 + 66 - 121 = 88$ and hence $U_2 = n_1 n_2 - U_1 = 11(13) - 121 = 55$

```
wilcox.test(SagebrushCrickets$time.to.mating ~ SagebrushCrickets$treatment, paired=FALSE)
       Wilcoxon rank sum test
data:  SagebrushCrickets$time.to.mating by SagebrushCrickets$treatment
W = 88, p-value = 0.3607
alternative hypothesis: true location shift is not equal to 0
```

Do not reject H_0; time to mating does not appear to be different for starved versus fed female crickets

What do we do if we have Tied Ranks?

The preceding example has no cases with the same rank (i.e., tied ranks)
If two observations have the same rank we give them the average rank $\left(\frac{i+j}{2}\right)$

If three observations have the same rank we give them the average rank $\left(\frac{i+j+k}{3}\right)$

... and so on

In the following example the question is whether blind versus sighted individuals use more or less gestures. These data have several ties as shown in the table on the following slide

1. The first 5 observations all have 0 gestures so the average rank becomes $\frac{1+2+3+4+5}{5}=3$
2. The next 12 observations are tied with 1 gesture so the average rank ends up being $\frac{6+7+8+9+10+\ldots+16+17}{12}=11.5$
3. The next 4 observations are tied with 2 gestures so the average rank ends up being $\frac{18+19+20+21}{4}=19.5$
4. The final 3 observations are tied with 3 gestures so the average rank ends up being $\frac{22+23+24}{3}=23$

Sightedness and Gestures

condition	gestures	rank	condition	gestures	rank
sighted	0	3	blind	1	11.5
sighted	0	3	blind	1	11.5
sighted	0	3	blind	1	11.5
blind	0	3	blind	1	11.5
blind	0	3	blind	1	11.5
sighted	1	11.5	sighted	2	19.5
sighted	1	11.5	sighted	2	19.5
sighted	1	11.5	sighted	2	19.5
sighted	1	11.5	blind	2	19.5
blind	1	11.5	sighted	3	23
blind	1	11.5	sighted	3	23
blind	1	11.5	blind	3	23

H_0: Number of gestures is not related to sightedness
H_A: Number of gestures is related to sightedness
Set $\alpha = 0.05$ and note that I am picking the *sighted* group for R_1 …

$$R_1 = 3+3+3+11.5+11.5+11.5+11.5+19.5+19.5+19.5+23+23 = 159.5$$

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 12(12) + \frac{12(13)}{2} - 159.5 = 62.5$$

$$U_2 = n_1 n_2 - U_1 = 144 - 62.5 = 81.5$$

```
> wilcox.exact(gestures$numberOfGestures ~ gestures$sightedness, paired=FALSE)
        Exact Wilcoxon rank sum test
data:  gestures$numberOfGestures by gestures$sightedness
W = 62.5, p-value = 0.6361
alternative hypothesis: true mu is not equal to 0
```

Do not reject H_0; number of gestures appear to be unrelated to sightedness

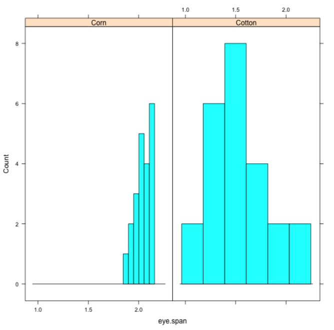
The Kolmogorov-Smirnov (K-S) test

- This (very weak) test is used to compare the distributions of two groups by comparing the empirical cumulative distribution functions (ecdfs) of the two groups and finding the greatest absolute distance between the two
- The $ecdf$ is $\hat{F}(Y) =$ fraction of sample with values $\leq Y_i$, where $i = 1, 2, 3, \cdots, n$
- The $K-S$ statistic is $D_{max} = |\hat{F}_1(Y) - \hat{F}_2(Y)|$
- Assumptions:
  1. The measurement scale is at least ordinal.
  2. The probability distributions are continuous
  3. The two samples are mutually independent
  4. Both samples are simple random samples from their respective populations
- H_0: $F_1(Y) = F_2(Y)$ for all Y_i
  H_A: $F_1(Y) \neq F_2(Y)$ for at least one Y_i
- Reject H_0 if $P-value$ of calculated $D_{max} \leq \alpha$

Eye to Eye

The stalk-eyed fly, *Crytodiopsis dalmanni*, is a bizarre-looking insect from the jungles of Malaysia. Its eyes are at the ends of long stalks that emerge from its head, making the fly look like something from the cantina scene in *Star Wars*. These eye stalks are present in both sexes, but they are particularly impressive in males. The span of the eye stalk in males enhances their attractiveness to females as well as their success in battles against other males. The span, in millimeters, from one eye to the other, was measured in a random sample of 45 male stalk-eyed flies.

The Stalkies' Example: Detailed K-S Test



- The graphical tests and the $var.test()$ tell us to reject both assumptions – the distributions may not be Normal and the variances may be unequal
- The $K-S$ test is ideally suited for this situation

Steps for the $K - S$ test

1. Arrange all eye.spans in ascending order
2. Calculate the cumulative frequency for each group – how many observations of group 1 have eye.spans $\leq$ each eye.span?
3. Convert these cumulative frequencies into relative cumulative frequencies (as proportions)
4. For each value of eye.span, calculate the difference between the relative cumulative frequency proportions of each group
5. Identify where the greatest difference is found
6. ... The logic? If the distributions are the same then the relative cumulative frequencies should be similar as well

Calculations for the Stalkies2 Example

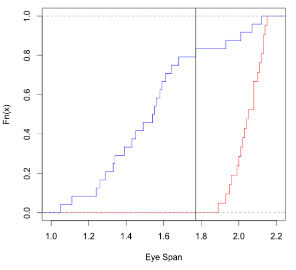
eye.span	cotton	corn	cotton.cf.p	corn.cf.p	diff	eye.span	cotton	corn	cotton.cf.p	corn.cf.p	diff
1.05	1	0	0.0417	0.0000	0.0417	1.89	20	1	0.8333	0.0476	0.7857
1.11	2	0	0.0833	0.0000	0.0833	1.93	21	2	0.8750	0.0952	0.7798
1.24	3	0	0.1250	0.0000	0.1250	1.95	21	3	0.8750	0.1429	0.7321
1.26	4	0	0.1667	0.0000	0.1667	1.96	21	4	0.8750	0.1905	0.6845
1.29	5	0	0.2083	0.0000	0.2083	1.99	21	5	0.8750	0.2381	0.6369
1.33	6	0	0.2500	0.0000	0.2500	2.00	21	6	0.8750	0.2857	0.5893
1.34	7	0	0.2917	0.0000	0.2917	2.01	22	7	0.9167	0.3333	0.5833
1.39	8	0	0.3333	0.0000	0.3333	2.02	22	8	0.9167	0.3810	0.5357
1.43	9	0	0.3750	0.0000	0.3750	2.03	22	9	0.9167	0.4286	0.4881
1.45	10	0	0.4167	0.0000	0.4167	2.04	22	10	0.9167	0.4762	0.4405
1.49	11	0	0.4583	0.0000	0.4583	2.05	22	11	0.9167	0.5238	0.3929
1.54	12	0	0.5000	0.0000	0.5000	2.07	23	11	0.9583	0.5238	0.4345
1.55	13	0	0.5417	0.0000	0.5417	2.08	23	14	0.9583	0.6667	0.2917
1.56	14	0	0.5833	0.0000	0.5833	2.10	23	15	0.9583	0.7143	0.2440
1.58	15	0	0.6250	0.0000	0.6250	2.11	23	16	0.9583	0.7619	0.1964
1.59	16	0	0.6667	0.0000	0.6667	2.12	24	17	1.0000	0.8095	0.1905
1.61	17	0	0.7083	0.0000	0.7083	2.13	24	19	1.0000	0.9048	0.0952
1.64	18	0	0.7500	0.0000	0.7500	2.14	24	20	1.0000	0.9524	0.0476
1.68	19	0	0.7917	0.0000	0.7917	2.15	24	21	1.0000	1.0000	0.0000
1.77	20	0	0.8333	0.0000	0.8333						

The cotton & corn columns list the cumulative frequencies of eye.span for each group

The columns ending in "cf.p" are these cumulative frequencies expressed as relative cumulative frequencies

diff is $= cotton.cf.p - corn.cf.p$ for each value of eye.span, and it is easy to see that the greatest difference occurs at $eye.span = 1.77$

The ecdf() and Test Results



```
ks.test(fed$eye.span, starved$eye.span, alternative="two.sided")

       Two-sample Kolmogorov-Smirnov test
data:  fed$eye.span and starved$eye.span
D = 0.8333, p-value = 3.51e-07
alternative hypothesis: two-sided
Warning message: In ks.test(fed$eye.span, starved$eye.span, alternative = "two.sided") : cannot compute correct p-values
       with ties
```

Reject H_0; diet has an effect on eye spans

Permutation Tests

- Useful alternatives to the Paired and the Two-sample t-test
- These tests generate a NULL distribution for the association between the numeric outcome variable and the categorical group variable
- Works by randomly reorganizing the original data (shown below)

Treatment	Time to mating	Treatment	Time to mating
starved	1.9	fed	1.5
starved	2.1	fed	1.7
starved	3.8	fed	2.4
starved	9.0	fed	3.6
starved	9.6	fed	5.7
starved	13.0	fed	22.6
starved	14.7	fed	22.8
starved	17.9	fed	39.0
starved	21.7	fed	54.4
starved	29.0	fed	72.1
starved	72.3	fed	73.6
		fed	79.5
		fed	88.9

(1) Randomly reassign Time to mating i

Treatment	Time to mating	Treatment	Time to mating
starved	3.8	fed	14.7
starved	9.0	fed	21.7
starved	3.6	fed	1.7
starved	79.5	fed	2.1
starved	17.9	fed	1.5
starved	22.8	fed	2.4
starved	54.4	fed	5.7
starved	13.0	fed	39.0
starved	9.6	fed	29.0
starved	1.9	fed	72.1
starved	22.6	fed	88.9
		fed	72.3
		fed	73.6

(2) Calculate $\bar{Y}_{starved} - \bar{Y}_{fed} = 21.65 - 32.67 = -11.02$

(3) Repeat this random juggling of the data at least 1,000 times, maybe even 10,000 times. Calculate $\bar{Y}_{starved} - \bar{Y}_{fed}$ for each permutation

(4) Calculate proportion of permutations with $\bar{Y}_{starved} - \bar{Y}_{fed} <= \bar{Y}_{starved} - \bar{Y}_{fed}$ in the original data (-18.26). This is the p-value (one-tailed)

```
> perm.test(cric$timeToMating ~ cric$feedingStatus, paired=FALSE, alternative="two.sided", exact
      =TRUE)
       2-sample Permutation Test (scores mapped into 1:(m+n) using rounded
       scores)
data:  cric$timeToMating by cric$feedingStatus
T = 123, p-value = 0.1376
alternative hypothesis: true mu is not equal to 0
```

Assumptions:

- Random samples
- Outcome is similarly shaped (i.e., distributed) in each group

Robust in large samples but low-powered (i.e., more likely to lead to Type II error) in small samples.

As powerful as t-tests in large samples

Testing Situations and Choices ...

- One-Sample or Paired Design
    1. Test for Normality
    2. If $N()$ holds then use the One-Sample $t-test$
    3. If $N()$ does not hold then try a transformation, test for $N()$ again
    4. If $N()$ still does not hold then use the Sign test (one sample) or the Wilcoxon Signed-Rank test (paired design)
    5. Permutation test
- Two Sample design (equal variances)
    1. Test for Normality and equal variances
    2. If $N()$ holds and variances are equal then use the Two-Sample $t-test$ with $var.equal = TRUE$
    3. If $N()$ does not hold then try a transformation, test for $N()$ again
    4. If $N()$ still does not hold then use the Mann-Whitney $U-test$
    5. Permutation test for Paired designs
- Two Sample design (unequal variances)
    1. Test for Normality and equal variances
    2. If $N()$ holds but variances are unequal then use the Two-Sample $t-test$ with $var.equal = FALSE$
    3. If $N()$ does not hold then try a transformation, test for $N()$ again
    4. If $N()$ still does not hold then use the Kolmogorov-Smirnov test
    5. Permutation test

Transformations and Back-transformations

- Popular Transformations
  - The Natural Logarithm: $Y^{'} = ln[Y]$ if $Y > 0$
  - The Natural Logarithm: $Y^{'} = ln[Y + 1]$ if $Y \geq 0$
  - The Square Root: $Y^{'} = \sqrt{Y + \frac{1}{2}}$
  - The Arcsine: $p^{'} = arcsin[\sqrt{p}]$
- The Corresponding Back-transformations
  - The Natural Logarithm: $Y = e^{Y^{'}}$
  - The Square Root: $Y = Y^{'2} - \frac{1}{2}$
  - The Arcsine: $p = \left(sin[p^{'}]\right)^2$