3 K-Sample Methods

General hypothesis test:

$H_0 : F_1(x) = F_2(x) = \ldots = F_k(x)$

H_a : For each (i, j): $F_i(x) \leq F_j(x),\ \forall x$, or $F_i(x) \geq F_j(x),\ \forall x$,

with strict inequality holding for at least one x and at least one pair (i, j).

Special case: The populations in H_a might differ only by location; i.e.,

$H_a : F_i(x) = F(x - \mu_i)$

In this special case of H_a, we may write

$$X_{ij} = \mu_i + \varepsilon_{ij},$$

where X_{ij} is the jth observation for the ith treatment, and the ε_{ij}s are independent and have the same distribution.

3.1.1 The F Statistic

Suppose $k \geq 2$ and the ε_{ij} are independent $N(0, \sigma)$.

We test for equality of all population means using an F-statistic, via *one-way analysis of variance* (ANOVA, from Math 321).

Let $\bar{X}$ be the overall mean.

Let $\bar{X}_i$, s_i, and n_i be the sample mean, sample standard deviation, and sample size, respectively, for the ith sample, for $i = 1, \ldots, k$.

Define the F-test statistic by:

$$F = \frac{\sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2/(k-1)}{\sum_{i=1}^{k}(n_i - 1)s_i^2/(N-k)} \quad \text{(need not memorize).}$$

Under H_0, the F-statistic has an F-distribution with $(k-1)$ degrees of freedom for the *numerator* and $(N-k)$ degrees of freedom for the *denominator*, where N is the total sample size.

Regarding power, do we prefer degrees of freedom to be large or small?

Graph the pdf for the following F-distributions: $F_{2,10}$, $F_{3,10}$, $F_{4,20}$.

Intuitively,

$$F = \frac{\text{variation between the treatment means, weighted by } n_i}{\text{variation among individuals within the same treatment}}$$

Do we prefer a *large* value of F or a *small* value of F?

At which tail would rejection of H_0 occur?

Problem 3.1.1 (corn and tobacco; quite hypothetical), problem3.1.1.txt: A farmer has *three* brands of fertilizer, and is testing if the different brands of fertilizer result in different mean yields of *corn*. Each fertilizer will be tested on *four* plots of land, so *twelve* similar plots of land will be involved in the experiment.

In a separate (independent) experiment, the *three* brands of fertilizer also are tested on a *tobacco* crop.

The data are given in the table below.

(a) Graph the data, and explain which crop, *corn* or *tobacco*, seems to be more influenced by the brand of fertilizer. Intuitively argue which crop should have the larger value of F.

Fertilizer	Corn yield	Tobacco yield
1	26	25
1	23	15
1	24	33
1	28	52
2	14	35
2	13	17
2	15	23
2	12	49
3	39	28
3	37	19
3	38	50
3	35	37

```
> z = read.table(
  "http://educ.jmu.edu/~garrenst/math324.dir/datasets/problem3.1.1.txt",
  header=TRUE )
> z = read.table2( "problem3.1.1.txt", header=TRUE )
> fertilizer = z[ , 1 ]
> corn = z[ , 2 ]
> tobacco = z[ , 3 ]
> plot( fertilizer, corn )
> plot( fertilizer, tobacco )
```

(b) How many degrees of freedom are associated with each test?

(c) Assume independent and identically distributed normal errors ε_{ij} in the model:

$$X_{ij} = \mu_i + \varepsilon_{ij},\ j = 1, \ldots, n_i;\ i = 1, 2, 3.$$

Test if the population mean *corn* yield is the same for all three brands of fertilizer versus the alternative that at least two of the mean *corn* yields are different.

```
> perm.f.test( corn, fertilizer, 0 )
```

(d) Under the assumption of normal errors, test if the mean *tobacco* yield is the same for all three brands of fertilizer versus the alternative that at least two of the mean *tobacco* yields are different.

```
> perm.f.test( tobacco, fertilizer, 0 )
```

□

Suppose the error terms (with positive finite standard deviation) are NOT normally distributed, but the sample size of each treatment is large. Is the approximation to the F-distribution still valid?

3.1.2 Steps in Carrying Out the Permutation *F* Statistic

For a *permutation* F-test, we use the same F-statistic, but we do not approximate the distribution of the F-statistic by an F-distribution.

Instead, p-values are based on the *permutation* distribution of the F-statistic.

To obtain this p-value:

⊙ Compute the value of F based on the *original* data.

⊙ Permute the *treatments* (or *responses*) to obtain a *permuted* F-statistic.

⊙ The *observed* F-statistic is compared to the values of the F-statistic under either *all* permutations or a *large number* of *simulated* permutations.

⊙ The p-value is the proportion of *permuted* F-statistics which are at least as **large** as the *observed* F-statistic.

How many groupings of the **treatments** are possible (for the sake of obtaining the **exact** p-value from the permutation F-test)?

Revisit Problem 3.1.1 (corn and tobacco), problem3.1.1.txt:

However, this time, do NOT assume that the three populations for crop yield are approximately normal.

How many groupings of the **treatments** are possible (for the sake of obtaining the **exact** p-value from the permutation F-test)?

(a) Test if the population mean *corn* yield is the same for all three brands of fertilizer versus the alternative that at least two of the mean *corn* yields are different.

(b) Test if the population mean *tobacco* yield is the same for all three brands of fertilizer versus the alternative that at least two of the mean *tobacco* yields are different.

□

F, chi-square, and normal distributions

Plot the χ^2_1, χ^2_2 and χ^2_3 probability density functions.

Plot the χ^2_4, χ^2_5 and χ^2_6 probability density functions.

Plot the χ^2_{10}, χ^2_{20} and χ^2_{30} probability density functions.

For large degrees of freedom, what distribution does a χ^2 distribution approximate?

◇

Recall that the ANOVA F-test has $(k-1)$ degrees of freedom for the *numerator* and

$(N-k)$ degrees of freedom for the *denominator*, where k is the number of *treatments* and N is the sample size.

Consider $k = 2$ treatments and sample sizes of $N = 3$ or $N = 4$ or $N = 5$.

Plot the $F_{1,1}$, $F_{1,2}$ and $F_{1,3}$ probability density functions.

For fixed $k = 2$ but large N, what distribution does an F-distribution approximate?

For any fixed k but large N, what distribution does $(k-1)F$ approximate, where F has an F-distribution with $(k-1)$ and $(N-k)$ degrees of freedom?

Homework p. 105: Exercises 3.1*, 3.2

Hints for homework exercise 3.1*: State H_0, H_a, and conclusion in statistical terms and in regular English, and **define** any notation used. Either retype your p-value as a comment using "#", or highlight the p-value in yellow. Remember to **COMPARE** your p-values. Introduce the question number as a comment using "#" or in red using .html code; e.g., `<span style="color: red"> Exercise 3.1 </span>`.

3.2 The Kruskal-Wallis Test

Are the *permutation* F-test and ANOVA F-test heavily influenced by outliers?

The *Kruskal-Wallis* test statistic is based on **ranks**.

Here, the populations are assumed to be identical, except for possibly location, and we test for equality of medians (or means, if finite).

Idea: Convert the original N observations to their appropriate **ranks**: 1, 2, 3, ..., N. Then, perform the *permutation* F-test using these **ranks** to obtain the *right*-tailed p-value. This p-value is the same one obtained when permuting the *Kruskal-Wallis* test statistic.

The *Kruskal-Wallis* test statistic is:

$$\text{KW} = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i \left(\bar{R}_i - \frac{N+1}{2} \right)^2 \quad \text{(need not memorize)},$$

where $\bar{R}_i$ is the average rank for the ith sample.

The modified formula for KW based on *ties* is given in section 3.2.2.

To obtain the *Kruskal-Wallis* p-value:

⊙ Compute the value of KW based on the *original* **ranks**.

⊙ Permute the *treatments* (or *ranks*) to obtain a *permuted* KW-statistic.

⊙ The *observed* KW-statistic is compared to the values of the KW-statistic under either *all* permutations or a *large number* of *simulated* permutations.

⊙ The p-value is the proportion of *permuted* KW-statistics at least as **large** as the *observed* KW-statistic.

The textbook shows that a strictly monotone increasing relationship exists between the *permutation* F-statistic based on **ranks** and the *Kruskal-Wallis* test statistic.

For large sample sizes (with fixed k) under H_0, the KW-test statistic is approximately χ^2 distributed, with how many degrees of freedom?

Is the *Kruskal-Wallis* test heavily influenced by outliers?

Revisit Problem 3.1.1 (corn and tobacco), problem3.1.1.txt:

(a) Read in the data, to perform tests based on the **TOBACCO** data.

```
> z = read.table2( "problem3.1.1.txt", header=TRUE )
> fertilizer = z[ , 1 ]


> tobacco = z[ , 3 ]
```

(b) Determine the value of the *Kruskal-Wallis* test statistic, without using the function `kruskal.test`.

```
> ranks = rank(tobacco)


> ni = c( 4, 4, 4 )
> N = sum( ni )

> mean.rank.i = c(mean(ranks[1:4]), mean(ranks[5:8]), mean(ranks[9:12]))


> KW = 12 / N / (N + 1) * sum( ni * ( mean.rank.i - (N + 1) / 2 ) ^ 2 )
```

(c) Approximate the p-value of the *Kruskal-Wallis* test, using the function `perm.f.test`.

(d) How many degrees of freedom are associated with the *Kruskal-Wallis* test?

(e) Determine the asymptotic p-value of the *Kruskal-Wallis* test statistic, without using the function `kruskal.test`.

(f) Obtain the *Kruskal-Wallis* statistic and p-value using `kruskal.test`.

```
> ?kruskal.test
```

□

Problem 3.2.1 (Birth conditions and IQ), problem3.2.1.txt: Steel (1959, *Biometrics*) presented the data below for testing whether certain conditions are associated with a lowering of IQ. The IQ score was obtained for 24 girls of which six each are *healthy*, *anoxic*, *premature*, and *Rh negative.*

Healthy	103	111	136	106	122	114
Anoxic	119	100	97	89	112	86
Rh negative	89	132	86	114	114	125
Premature	92	114	86	119	131	94

(a) Graph the data. Intuitively, do the four populations of birth conditions seem to differ regarding IQ?

```
> IQ = scan2( "problem3.2.1.txt" ) # Read in data as a vector.
> birth.condition = rep( 1:4, each=6 )

> plot( birth.condition, IQ )
```

(b) Assume independent and identically distributed errors ε_{ij} with mean zero in the model:

$$X_{ij} = \mu_i + \varepsilon_{ij},\ j = 1, \ldots, n_i;\ i = 1, \ldots, 4.$$

Formulate the null and alternative hypotheses for testing if the mean *IQ* scores are the same for all four types of *birth conditions* versus the alternative that at least two of the mean *IQ* scores are different.

(c) How many degrees of freedom are associated with the *Kruskal-Wallis* test?

(d) Perform the *Kruskal-Wallis* test.

(e) How many degrees of freedom are associated with the ANOVA F-test?

(f) Perform the ANOVA F-test.

(g) Perform the *permutation* F-test.

(h) For which test(s), *Kruskal-Wallis* test, ANOVA F-test, or *permutation* F-test, is the assumption of **normal** errors ε_{ij} **least** relevant?

□

Homework p. 105: Exercise 3.3*

Hints for homework exercise 3.3*: State H_0, H_a, and conclusion in statistical terms and in regular English, and **define** any notation used. Either retype your p-value as a comment using "#", or highlight the p-value in yellow. Remember to **COMPARE** your p-values. Introduce the question number as a comment using "#" or in red using .html code; e.g., `<span style="color: red"> Exercise 3.3 </span>`.

ercises

1 The data are samples from three simulated distributions.

Group 1	2.9736	0.9448	1.6394	0.0389	1.2958
Group 2	0.7681	0.8027	0.2156	0.074	1.5076
Group 3	4.8249	2.2516	1.5609	2.0452	1.0959

a Apply the permutation F-test to the data.

b Compare the results in part a with the results of the usual one-way analysis of variance.

2 The following data from the National Transportation Safety Administration are the left femur loads on driver-side crash dummies for automobiles in various weight classes. Apply the permutation F-test and the ANOVA F-test to the data, and compare p-values. Does it appear that the data are normally distributed? The complete data set may be obtained from the Data and Story Library at http://lib.stat.cmu.edu/DASL/.

Vehicle Weight Classification				
1700 lb	*2300 lb*	*2800 lb*	*3200 lb*	*3700 lb*
574	791	865	998	1154
976	1146	775	1049	541
789	394	729	736	406
805	767	1721	782	1529
361	1385	1113	730	1132
529	1021	820	742	767
	2073	1613	1219	1224
	803	1404	705	314
	1263	1201	1260	1728
	1016	205	611	
	1101	1380	1350	
	945	580	1657	
	139	1803	1143	

3 Refer to the data in Exercise 1. Apply the Kruskal–Wallis test to the data, and compare the conclusions with those obtained in Exercise 1.

4 Obtain the permutation distribution of the Kruskal–Wallis statistic when $n_1 = n_2 = n_3 = 2$.

5 An agronomist gave scores from 0 to 5 to denote insect damage to wheat plants that were treated with four insecticides. The data are given in the following table. Use the Kruskal–Wallis test with ties to test whether or not there is a difference among the treatments.