

# Nonparametric Methods: Permutation Tests

BIOS 6611

CU Anschutz

Week 6

1 Permutation Tests

2 Permutation Test for Independence of Two Variables

# Permutation Tests

# Motivation

Up until now we have mostly considered parametric procedures, i.e. those that, based on the central limit theorem, depend on the normality of the statistics being computed. Nonparametric methods are useful when the assumption of normality does not hold (e.g. small samples, heavily skewed data, ordinal data).

# Permutation Tests

**Permutation tests** are also known as randomization tests and exact tests. In permutation testing we randomly “exchange labels” on data points to generate a null, or permutation, distribution. From this we are able to calculate a p-value to summarize how extreme our sample appears to be.

In this case, we will sample without replacement from a pooled sample of all data, and each observation will only be represented once in each permutation resample.

However, now the group membership (or other features) will vary based on resampling, so our previous observation from Group 1 in our original sample may be in either Group 1 or Group 2 for any given permutation sample.

# Comparison to Bootstrap Sampling

- Permutation tests sample *without* replacement, bootstrap sampling samples *with* replacement
- Permutation distributions are centered at the null hypothesis value, the bootstrap distribution is approximately centered at the original mean value (difference, proportion, SD, etc.) from the data
- Permutation tests result in a p-value, bootstrap samples result in an estimate of the SE of the estimators and allow us to combine confidence intervals
- It is inappropriate to combine them together for permutation p-value and a bootstrap CI due to the different assumptions and sampling strategies

# Two-Sample Permutation Test

Pool the  $m$  values from Sample 1 and  $n$  values from Sample 2.

**Repeat the following steps:**

- 1 Draw a resample of size  $m$  without replacement to represent Sample 1.
- 2 Use the remaining  $n$  observations to represent Sample 2.
- 3 Calculate the difference in means (or any other statistic to compare the samples).

**until we have “enough” samples.**

Calculate the p-value as the fraction of times the random statistics exceed or are equal to the original statistic. Multiply by 2 for a two-sided test.

Optionally, plot a histogram of the random statistic values.

# Permutation Test Assumptions and Notes

The distribution of your difference (or other statistic) across all permutation resamples is the **permutation distribution**. It can be exact (exhaustively calculate all permutations) or approximate (implemented with sampling).

For permutation testing, the distribution doesn't have to be normal. It can be anything, so long as the two populations (groups) have the same distribution under the null hypothesis. Thus, group labels are said to be **exchangeable**. Differences in distributions under  $H_0$  can yield misleading results.

The summary statistic for each permutation can be a difference in means, medians, proportions, a test statistic, etc. so the approach can be easily generalized. For example, permutation tests have been described for regression models.



## Example 1 - Comparing Group Means

Example 5.4 from our *Chihara and Hesterberg* textbook is a comparison of commercial length between basic and extended cable during random half-hour periods from 7am-11pm:

Basic	7.0	10.0	10.6	10.2	8.6	7.6	8.2	10.4	11.0	8.5
Extended	3.4	7.8	9.4	4.7	5.4	7.6	5.0	8.0	7.8	9.6

We will compare the difference in means between the two cable types and calculate the permutation test p-value.

# Implement the Permutation Test

```
set.seed(515)

times.Basic <- c(7,10,10.6,10.2,8.6,7.6,8.2,10.4,11.0,8.5)
times.Ext <- c(3.4,7.8,9.4,4.7,5.4,7.6,5.0,8.0,7.8,9.6)
observed_diff <- mean(times.Basic) - mean(times.Ext)

pool_dat <- c(times.Basic, times.Ext) # Combine data into one vector

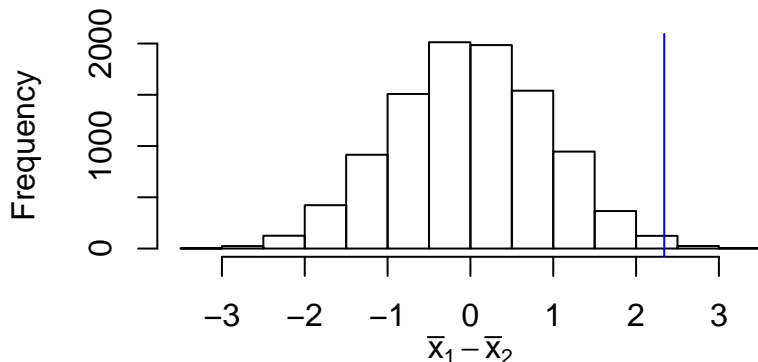
N <- 104 - 1 # number of permutations
result <- numeric(N)

for(i in 1:N){
  # Sample 10 values without replacement
  index <- sample(length(pool_dat), size=10, replace=FALSE)
  # Calculate difference
  result[i] <- mean(pool_dat[index]) - mean(pool_dat[-index])
}
```

# Plot the Permutation Distribution

```
hist( result, xlab='',main='Permutation Dist for Diff in Commercials')  
mtext(text=expression(bar(x)[1]-bar(x)[2]), side=1, line=2)  
abline(v = observed_diff, col='blue')
```

## Permutation Dist for Diff in Commercials



# Calculate the Permutation Test p-value

We can calculate a two-sided p-value by taking the larger one-sided estimate and multiplying by 2:

```
(sum(result >= observed_diff) + 1)/(N+1)
## [1] 0.0056
(sum(result <= -observed_diff) + 1)/(N+1)
## [1] 0.0048
```

$p = 0.0056 \times 2 = 0.0112 < 0.05$ . This suggests there is a significant difference in commercial length between basic and extended cable.

Note that we add 1 to the numerator and denominator. This represents the original data as an extra resample (i.e., there is always at least one resample that is as extreme as the original data).

Here we can also more directly see the definition of a p-value as *observing something as or more extreme than our given sample assuming the null hypothesis is true*.

# Permutation Test for Independence of Two Variables

## Motivating Example

Survey data from 2001 on support for marijuana for medicinal purposes: Does support for medical marijuana depend on age?  $H_0$ : no association between age and favoring use of medical marijuana.

Age Group	For	Against
18-29 yo	172	52
30-49 yo	313	103
>50 yo	258	119

We could evaluate this with a chi-squared test since it is a contingency table.

However, perhaps we are concerned about making assumptions of the asymptotic distribution. Instead, we can conduct a permutation test on the  $X^2$  itself without assuming the  $\chi^2$  distribution.

# Permutation Test for Independence of Two Variables

Store the data in a table with one row per observation and one column per variable.

Calculate a test statistic for the original data. Normally large values of the test statistic suggest dependence.

**Repeat the following steps:**

- 1 Randomly permute the rows in one of the columns (i.e., hold the other fixed).
- 2 Calculate the test statistic for the permuted data.

**until we have “enough” samples.**

Calculate the p-value as the fraction of times the random statistics exceed or are equal to the original statistic.

Optionally, plot a histogram of the random statistic values.

# Permutation Test for Independence - I

```
X <- matrix(nrow=3, byrow=T, c(172,52,313,103,258,119),  
           dimnames=list(c("18-29", "30-49", ">50"), c("for", "against")))  
  
prop.table(X, margin=1)
```

```
##           for   against  
## 18-29 0.7678571 0.2321429  
## 30-49 0.7524038 0.2475962  
## >50   0.6843501 0.3156499
```

```
library(epitools) # for expand.table()  
dat <- expand.table(X) # transform data from 3x2 table to 1 obs per row  
  
# define a function to do the chisquare test  
chisq <- function(Obs){  
#Obs is the observed contingency table  
  Expected <- outer(rowSums(Obs), colSums(Obs))/sum(Obs)  
  sum((Obs-Expected)^2/Expected)  
}
```



# Permutation Test for Independence - II

```
agegrp <- dat[,1]
response <- dat[,2]
observed <- chisq(table(agegrp, response))
observed
```

```
## [1] 6.681429
```

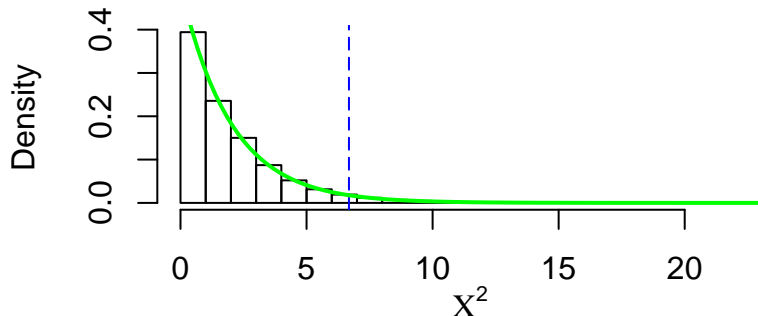
```
N <- 105-1 #set number of times to repeat this process
result <- numeric(N) # space to save the random differences
set.seed(515)
```

```
for(i in 1:N){
  agegrp.permuted <- sample(agegrp)
  perm.table <- table(agegrp.permuted, response)
  result[i] <- chisq(perm.table)
}
```

# Permutation Test for Independence - III

```
hist(result, freq=FALSE, xlab='', main="Permutation dist for chi-square stat",  
mtext(expression(Chi^2), side=1, line=2)  
abline(v = observed, col = "blue", lty=5)  
curve(dchisq(x, 2), add=TRUE, col="green", lwd=2)
```

## Permutation dist for chi-square statistic



# Permutation Test for Independence - IV

Now let's calculate the p-value for our test of independence:

```
(sum(result >= observed)+1)/(N + 1) #P-value
```

```
## [1] 0.03591
```

$p=0.03591 < 0.05$ , so we would reject  $H_0$  that there is no association between age and favoring the use of medical marijuana.

We can also compare to the p-value from a chi-squared test:

```
1-pchisq(observed, df=2) # compare to chi-squared dist.
```

```
## [1] 0.03541165
```

```
chisq.test(X) # p-value directly from chisq.test
```

```
##
```

```
## Pearson's Chi-squared test
```

```
##
```

```
## data: X
```

```
## X-squared = 6.6814, df = 2, p-value = 0.03541
```

# Final Permutation Testing Notes

- Various statistics can be used: e.g. mean, median, proportions, etc.
- For more precision, a larger number of permutations should be used.
- Sampling (entire) permutations without replacement is most appropriate but with replacement is acceptable and faster (i.e., we don't necessarily need to check that a given permutation sample is unique amongst all our previous permutation samples).
- Normality of underlying distributions not assumed. Robust to skewness and imbalance as long as the underlying distributions are equal under the null hypothesis.
- No random sampling assumption is required. If random sampling is not assumed, inference to a population can't then be made but a conclusion about the sample can be drawn. However, treatment (exposure) assignments are assumed to be random.