STAT 210
Applied Statistics and Data Analysis
Experimental Design II

Joaquin Ortega

```r
library(PASWR)
```

```
## Loading required package: e1071
```

```
## Loading required package: MASS
```

```
## Loading required package: lattice
```

```r
mod0 <- lm(StopDist ~ tire, data = Tire)
```

Power and the number of replicates

Power and the number of replicates

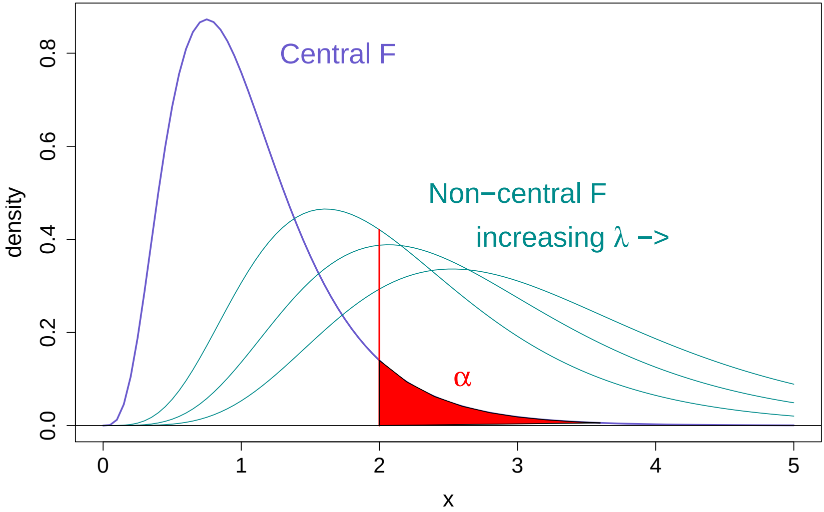
The test statistic msT/msE follows the F-distribution when the null hypothesis is true, but when the null hypothesis is false it follows the noncentral F-distribution.

The noncentral F-distribution has a wider spread than the central F-distribution.

The spread in the noncentral F-distribution and probability exceeding the critical limit from the central F-distribution is an increasing function of the noncentrality parameter, λ.

When the distribution is the noncentral F, the probability of exceeding the critical limit from the central F-distribution is the power of the test.

Power and the number of replicates



Power and the number of replicates

The power can be computed for any scenario of differing means, if the values of the cell means, the variance of the experimental error, and the number of replicates per factor level are specified.

For a constant difference among cell means, represented by $\sum_i(\mu_i - \mu)^2$, the noncentrality parameter and the power increase as the number of replicates increase.

When the differences among cell means is large enough to have practical importance, the experimenter would like to have high power, or probability of rejecting the hypothesis of no treatment effects.

When the difference among the means has practical importance to the researcher we call it **practical significance**.

Practical significance does not always correspond to statistical significance as determined by the F-test from the ANOVA.

Power and the number of replicates

If there is a difference among the cell means, the power is given by

$$Power(\lambda) = \int_{F_{k-1,k(r-1)\alpha}}^{\infty} f(x, k-1, k(r-1), \lambda)\, dx$$

where

- $F_{k-1,k(r-1)\alpha}$ is the α-th percentile of the central F distribution, with $k-1$ and $k(r-1)$ degrees of freedom
- $f(x, k-1, k(r-1), \lambda)$ is the non-central F density with non-centrality parameter λ and
- $\lambda = \frac{r}{\sigma^2} \sum_{i=1}^{k} (\mu_i - \bar{\mu}_\bullet)^2$

For a fixed value of $\frac{1}{\sigma^2} \sum_{i=1}^{k} (\mu_i - \bar{\mu}_\bullet)^2$ the power increases with r.

These computations can be carried out with the `Fpower1` function in the `daewr` package.

Power and the number of replicates

In the tire example suppose that the standard tread is `D` and a difference of less than 30 feet in the braking distance is of no interest to the manufacturer, but a difference larger than this value would be of interest.

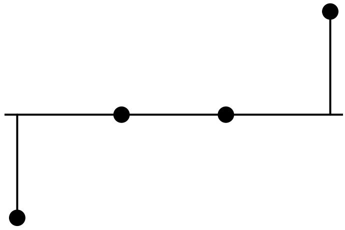
In this case we regard $\Delta = 30$ as a practical difference in cell means.

We need a value, or at least a lower bound for the sum $\sum_{i=1}^{k}(\mu_i - \bar{\mu}_{\bullet})^2$ under the condition that at least one of the values for the stoping distance for treads `A`, `B` or `C` differ from `D` by at least 30 feet.

Power and the number of replicates

The minimum value is attained when the result for one of the cell means for A,B or C is lower than the average by $\Delta/2$, the braking distance for D is higher than the average by $\Delta/2$ and the other two values are equal to the average.

```
plot(1:4,c(-1,0,0,1), pch=19, axes = FALSE, xlab='',
     ylab='')
abline(h=0); segments(1,0,1,-1); segments(4,0,4,1)
```



Power and the number of replicates

This results in

$$\sum_{i=1}^{k}(\mu_i - \bar{\mu}_\bullet)^2 = \left(\frac{\Delta}{2}\right)^2 + 0 + 0 + \left(\frac{\Delta}{2}\right)^2 = \frac{\Delta^2}{2} = 450.$$

By previous experience the manufacturer knows that a reasonable estimate for the variance of the braking distance is 225 ft^2. The noncentrality parameter can be calculated as

$$\lambda = \frac{r}{225}450.$$

The power is calculated for $r = 2, \ldots, 10$ using the `Fpower1` function in the `daewr` package.

Power and the number of replicates

```r
library(daewr)
rmin <-2 #smallest number of replicates considered
rmax <-10 # largest number of replicates considered
alpha <- rep(0.05, rmax - rmin +1)
sigma <- 15; nlev <- 4; nreps <- rmin:rmax; Delta <- 30
(power <- Fpower1(alpha,nlev,nreps,Delta,sigma))
```

```
##       alpha nlev nreps Delta sigma     power
##  [1,]  0.05    4     2    30    15 0.1698028
##  [2,]  0.05    4     3    30    15 0.3390584
##  [3,]  0.05    4     4    30    15 0.5037050
##  [4,]  0.05    4     5    30    15 0.6442332
##  [5,]  0.05    4     6    30    15 0.7545861
##  [6,]  0.05    4     7    30    15 0.8361289
##  [7,]  0.05    4     8    30    15 0.8935978
##  [8,]  0.05    4     9    30    15 0.9325774
##  [9,]  0.05    4    10    30    15 0.9581855
```

Power and the number of replicates

Non-Parametric Tests

Non-Parametric Tests

Non-parametric procedures make no distributional assumptions about the data being analyzed. In this sense, they free us from the possible pitfalls of choosing a wrong distribution.

The price to pay is usually having less power associated to these procedures, when the hypotheses upon which the parametric methods are based are true, which seems natural since non-parametric procedures they will work for general samples.

Comparing Two Populations: The Rank Sum Test

We begin by introducing the rank sums test, which was proposed by Wilcoxon in 1945. Later, Mann and Whitney proposed another test, which is equivalent to Wilcoxon's. Therefore, any combination of these three surnames may appear to refer to this test.

In this test we are interested in differences in means. Assume that we have two samples $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ from two continuous symmetric distributions with the same variance but (possibly) different mean. If the distributions are not symmetric, a test of medians is possible.

Comparing Two Populations: The Rank Sum Test

If the distributions are the same, we can consider that the pooled sample $x_1, \ldots, x_n, y_1, \ldots, y_m$ is a random sample of size $N = n + m$ from the common distribution. Hence, if these observations are ordered according to magnitude, we would expect to see the xs and ys well mixed. If one of these sets tends to cluster together at some extreme of the ordered sample, we would consider this as evidence that the null hypothesis of equal distributions is not true.

Comparing Two Populations: The Rank Sum Test

Let μ_X be the (population) mean for the X distribution and similarly for μ_Y. Our test is

$$H_0 : \mu_X = \mu_Y \qquad \text{vs.} \qquad H_A : \mu_X \neq \mu_Y .$$

The (Wilcoxon) rank sums test proceeds as follows: The two samples are joined together giving a sample of size $N = n + m$. This sample is ordered and the ranks (positions in the ordered sample) for the elements of the x sample are added up. This is the test statistic W. If the null hypothesis of equal distributions is true, all $\binom{N}{n}$ possible assignments of ranks for the x sample are equally likely, each having probability $1/\binom{N}{n}$.

Comparing Two Populations: The Rank Sum Test

When the samples have equal sizes $n = n_1 = n_2$, the test statitic W takes integer values ranging from $n(n+1)/2$ to $n(2N - n + 1)/2$ when there are no ties in the ranks. The distribution of W is known as Wilcoxon rank-sum distribution and it can be otained in `R` with the command `pwilcox`.

Example

Two samples of fish were drawn from different lakes and the fish were weighted. The weights in grams are

```
sampl1 <- c(286, 251, 325, 313, 309, 308)
sampl2 <- c(249, 324, 289, 303, 310, 318)
```

we build a data frame with this information

```
sample <- c(rep(1,6), rep(2,6))
wcx <- data.frame(weight=c(sampl1,sampl2), sample)
str(wcx)
```

```
## 'data.frame':    12 obs. of  2 variables:
##  $ weight: num  286 251 325 313 309 308 249 324 289 303
##  $ sample: num  1 1 1 1 1 1 2 2 2 2 ...
```

Comparing Two Populations: The Rank Sum Test

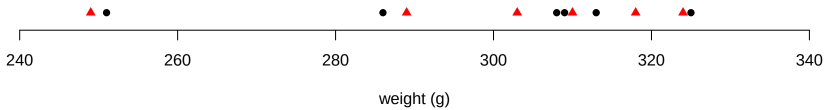
We can order the results by weight using the `order` function. We also add a new column with the rank.

```
ord <- order(wcx[,1])
wcx.ord <- cbind(wcx[ord,],rank=1:12)
head(wcx.ord)
```

```
##    weight sample rank
## 7     249      2    1
## 2     251      1    2
## 1     286      1    3
## 9     289      2    4
## 10    303      2    5
## 6     308      1    6
```

Comparing Two Populations: The Rank Sum Test

We plot the ordered values



Comparing Two Populations: The Rank Sum Test

Next, we sum the ranks for the observations in the pooled data:

```
(rank.sum <- c( sum(wcx.ord[,3][wcx.ord[,2]==1]),
                sum(wcx.ord[,3][wcx.ord[,2]==2]) ))
```

```
## [1] 39 39
```

In this example the rank sums are equal, and this is strong evidence in favour of the null hypothesis of equal means.

Comparing Two Populations: The Rank Sum Test

The command `wilcox.test()` in R performs the rank sums test:

```
wilcox.test(sampl1,sampl2)
```

```
##
##  Wilcoxon rank sum test
##
## data:  sampl1 and sampl2
## W = 18, p-value = 1
## alternative hypothesis: true location shift is not equal
```

Comparing Two Populations: The Rank Sum Test

Comparing with the t-test, which assumes normality, the rank sums test has less power than the t-test if the distribution is normal, but if departures from normality in the sample are marked, the t test is inferior and the p-values cannot be trusted.

The Kuskal-Wallis test

The Kruskal-Wallis test is an extension of the rank sum test to the case of d multiple samples.

The null hypothesis is that all the samples come from the same distribution while the alternative is that at least two of the samples come from different distributions.

The only requisite is that the population distributions have to be continuous.

The Kuskal-Wallis test

Because the underlying distributions of the d populations are assumed to be identical under the null hypothesis, this test can be applied to means, medians, or any other quantile.

The null and alternative hypotheses are expressed in terms of the means as

$$H_0 : \mu_1 = \cdots = \mu_d \quad \text{vs.} \quad H_A : \mu_i \neq \mu_j \text{ for at least one pair of } i, j.$$

To test the null hypothesis the $n_1, n_2, \ldots, n_d$ observations are pooled together and ordered from 1 to $N = n_1 + \cdots + n_d$ to obtain the ranks.

The Kuskal-Wallis test

The standardized test statistic used by `R` is

$$H = \frac{12 \sum_{i=1}^{d} n_i(\bar{R}_i - \bar{R})}{N(N+1)},$$

where

- n_i is the number of observations for the i-th treatment,
- $\bar{R}_i$ is the average of the ranks in the i-th treatment and
- $\bar{R}$ is the overall average of the ranks.

The Kuskal-Wallis test

When there are ties in the average ranks for the groups, adjustments in the test statistic must be made.

As the size of the smallest group goes to infinity, the test statitic converges in distribution to a χ^2 distribution with $d - 1$ degrees of freedom.

To use this approximation it is usually required that the minimim group size be at least five.

In `R`, `kruskal.test()` performs this test with the corresponding corrections when ties are present.

Let us compare the results with this test to those obtained before for the `yields` example.

The Kuskal-Wallis test

We first reproduce the Anova table and then we do the K-W test:

```
results <- read.table('yields.txt',header=T)
frame <- stack(results)
names(frame) <- c('yield','soil')
with(frame,summary(aov(yield~soil)))
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## soil         2   99.2   49.60   4.245  0.025 *
## Residuals   27  315.5   11.69
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

The Kuskal-Wallis test

```
with(frame, kruskal.test(yield~soil))
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  yield by soil
## Kruskal-Wallis chi-squared = 7.5813, df = 2, p-value = 0
```

We get similar p-values in this case, 0.025 in the Anova table and 0.0226 in the K-W test, so we would reach the same conclusion for the usual values of α.

The Kuskal-Wallis test

The other example we considered was the tire experiment:

```r
anova(mod0)
```

```
## Analysis of Variance Table
##
## Response: StopDist
##           Df Sum Sq Mean Sq F value   Pr(>F)
## tire       3 5673.1 1891.04  5.3278 0.007316 **
## Residuals 20 7098.8  354.94
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Kuskal-Wallis test

```r
with(Tire, kruskal.test(StopDist ~ tire))
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  StopDist by tire
## Kruskal-Wallis chi-squared = 9.9133, df = 3, p-value = (
```

In this case results are not so close. p values are 0.007 and 0.019, so at the 1% level we would reach reach different conclusions.