• We now consider an analysis with only categorical predictors (i.e. all predictors are factors).
  – Predicting height from sex (M/F)
  – Predicting white blood cell count from treatment group (A,B,C)

• If only 1 categorical predictor of a continuous response ⇒ One-Way ANOVA

  \[ \begin{array}{ccc} 
  \mu_1 & \mu_2 & \mu_3 \\
  \end{array} \]

  For example,

  \[ \begin{array}{ccc} 
  \mu_{dem} & \mu_{rep} & \mu_{ind} \\
  \end{array} \]

1-way ANOVA (only one factor)

• Consider the cell means model notation: (It uses 1 parameter to represent each cell mean)

  \[ Y_{ij} = \mu_i + \epsilon_{ij} \]

  where \( \mu_i \) is the group \( i \) mean.

• If there’s only 2 levels, like in sex, then we can use a two-sample \( t \)-test \( H_0 : \mu_1 = \mu_2 \).

• If we have more than 2 levels, we extend this \( t \)-test idea to do a 1-way ANOVA.

• A two-sample \( t \)-test is essentially a 1-way ANOVA (it’s the simplest one, there’s only 2 levels to the factor)

1-way ANOVA visual:

In a 1-way ANOVA, we’re interested in finding differences between population group means, if they exist.

• If two categorical predictors of a continuous response ⇒ Two-Way ANOVA

  \[ \begin{array}{ccc} 
  \mu_{11} & \mu_{12} & \mu_{13} \\
  \mu_{21} & \mu_{22} & \mu_{23} \\
  \end{array} \]

  For example,

  \begin{array}{ccc}
  \text{Factor 1} & \text{low} & \text{med} & \text{high} \\
  \text{Factor 2} & \text{yes} & \text{no} \\
  \end{array}

• Suppose we have three populations (or 3 levels of a categorical variable) to compare...

Example: Does the presence of pets or friends affect the response to stress?

\( n = 45 \) women (all dog lovers)

Each woman randomly assigned to one of three treatment groups as:

1) alone
2) with friend
3) with pet

Their heart rate is taken and recorded during a stressful task.

> pets=read.csv("pets.csv")
> head(pets)
group  rate
1   P 69.169
2   F 99.692
3   P 70.169
4   C 80.369
5   C 87.446
6   P 75.985
> attach(pets)
> is.factor(group)
[1] TRUE

The treatment groups are:
'C' for control group or *alone.*
'F' for *with friend.*
'P' for *with pet.*

Consider the distribution of heart rate by treatment group...

Get the mean of each group.
> tapply(rate,group,mean)
       C     F     P
82.5207 91.3251 73.48307

If we consider $\mu_1$ as the population mean heart rate of the control group,
$\mu_2$ as the population mean heart rate of the friends group,
$\mu_3$ as the population mean heart rate of the pet group,

then, to test if any of the groups have a different heart rate, we would consider the ‘overall’ null hypothesis

$H_0: \mu_1 = \mu_2 = \mu_3$
$H_A$: at least one $\mu_i$ is different for $i=1,2,3$

Why not just do 3 pairwise comparisons?

- $H_0: \mu_1 = \mu_2$
- $H_0: \mu_1 = \mu_3$
- $H_0: \mu_2 = \mu_3$

Problems:

- 3 separate p-values for 3 different tests don’t tell us how likely it is that *three* sample means ($\bar{Y}_1, \bar{Y}_2, \bar{Y}_3$) are spread apart as far as these by chance (i.e. when $\mu_1 = \mu_2 = \mu_3$).

- It might be that $\bar{Y}_1 = 73.48$ and $\bar{Y}_2 = 91.32$ are significantly different when only looking at 2 populations, but *not* significantly different if we look at 3 populations
  - as more and more groups are added, we expect the gap between smallest and largest sample means to get larger (even if the population means are the same).
• The probability of a Type I error (rejecting $H_0$ when it was true) for the whole set of three $t$-tests will be larger than $\alpha$.

  – For example,
    * Set the chance of making a Type I error on an individual $t$-test at $\alpha = 0.05$
    * The chance of NOT making a mistake on an individual $t$-test is $(1 - \alpha) = 0.95$

  * What is the probability of not making a mistake on any of the three tests?

    Well, if the tests were independent,
    $$(1 - \alpha)(1 - \alpha)(1 - \alpha) = (0.95)^3 = 0.8574$$

    Thus, there’s a 0.1426 chance of making at least 1 mistake (definitely larger than $\alpha = 0.05$ for the whole set of three tests)

Multiple comparisons procedures in statistics

• How do we do numerous comparisons of interest simultaneously while maintaining a certain overall error rate (like $\alpha = 0.05$)?

• Two steps
  – start with an overall test to see if anything is interesting... i.e. test if any of the means are significantly different using an overall $F$-test
  – if so, do a follow up analysis to decide which groups differ and to estimate the size of differences

• Step one: 1-Way ANOVA $F$-test
  We perform the overall $F$-test:
  $H_0 : \mu_1 = \mu_2 = \mu_3$
  $H_A : \text{at least one } \mu_i \text{ is different for } i=1,2,3$

  Though we can do coding of two dummy variables to represent group, R will do this for us.

  > is.factor(group)
  [1] TRUE

  # See how R has coded the dummy variables for
  # the 'factor' variable group using contrasts():
  > contrasts(group)
  F P
  C 0 0
  F 1 0
  P 0 1

  So, C is the baseline group (because both dummy variables are 0).
The regression model with dummy variables:

$$Y_i = \beta_0 + \beta_F D_{1i} + \beta_P D_{2i} + \epsilon_i$$

The model for each group (3 separate means):

- Control group: $Y_i = \beta_0 + \epsilon_i$
- Friend group: $Y_i = (\beta_0 + \beta_F) + \epsilon_i$
- Pet group: $Y_i = (\beta_0 + \beta_P) + \epsilon_i$

The expected value of $Y$ is the same for all $i$ within a group.

The dummy variable parameters represent the difference between the respective group and the baseline group.

$$\hat{\beta}_F = \text{Friend group mean - Baseline group mean}$$

$$\hat{\beta}_P = \text{Pet group mean - Baseline group mean}$$

Here, baseline was the Control group.

We can use the 'anova' function to get the sums of squares for the model (or the group variable), the residual sums of squares, and the overall F-test.

```
> anova(lm.out)  ## Only 1 factor used in this model
Analysis of Variance Table

Response: rate
       Df Sum Sq Mean Sq F value Pr(>F)
group   2 2387.7 1193.89 14.080 2.09e-05 ***
Residuals 42 3561.3  84.80
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
```

The summary statement also gives us the overall F-test, AND the tests for $\beta_F = 0$, and $\beta_P = 0$.

```
> summary(lm.out)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   82.524      2.378 34.709  < 2e-16 ***
groupF        8.801      3.362  2.617   0.012 *
groupP      -9.041      3.362 -2.689   0.010 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
```

Rejection of $H_0: \beta_F = 0$ says there is evidence that the Friend group is different from the baseline group.

Rejection of $H_0: \beta_P = 0$ says there is evidence that the Pet group is different from the baseline group.

The overall F-test tests $H_0: \beta_F = \beta_P = 0$, which is the same as $H_0: \mu_C = \mu_F = \mu_P$. 

What do we get when we fit the model in R.

```
> is.factor(group)
[1] TRUE

> lm.out=lm(rate ~ group)

> coefficients(lm.out)
         (Intercept)  groupF  groupP
82.524067     8.801067   -9.041000

The group means:

> tapply(rate,group,mean)
   group   
    C     82.52407  
    F     91.32513  
    P     73.48307 

The distributions:

14
General concept of 1-Way ANOVA

What matters is how far apart the sample means are relative to variability.

The means of the three groups above is the same in both pictures, but the variability is quite different.

Roughly speaking, we’ll consider an $F$-statistic where...

$$F = \frac{\text{variation among the sample means}}{\text{variation among individuals in the same sample}}$$

- When there is a lot of spread between the group sample means and very little spread within a group, we will have a large $F$-statistic.

- Compare $F$ to an $F$-distribution for decision

Notation:

- $m$ = number of different populations whose means we are studying
- $n_i$ = number of observations in sample from $i$th population
- $N = \text{total } # \text{ of observations}, N = \sum_{i=1}^{m} n_i$

$F$-distribution for 1-way ANOVA $F$-test

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_m$

$H_A : \text{not } H_0$

- $m-1$ numerator degrees of freedom
- $N-m$ denominator degrees of freedom

($N$ data points, $m$ parameters estimated)

Assumptions for One-way ANOVA

- $m$ independent simple random samples, one from each of $m$ populations

- Each population $i$ is normally distributed about its unknown mean $\mu_i$
  - If sample size is large enough, the Central Limit Theorem will kick-in and inferences based on sample means will be OK even if the populations distributions are not exactly normal.
  - If you don’t have normality, you could use a non-parametric test, such the Kruskall-Wallis test which is based on the ranks of the $y$-values rather than the $y$-values themselves.

- The populations have the same variance, $\sigma^2$
  - You can use Levene’s Test (in the car library) to test for non-constant variance.
\[ H_0 : \sigma_1 = \sigma_2 = \sigma_3 \]

\[
\text{> leveneTest(rate, group)}
\]

Levene’s Test for Homogeneity of Variance

<table>
<thead>
<tr>
<th></th>
<th>DF</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>2</td>
<td>0.0028</td>
<td>0.9973</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Since the p-value is not less than 0.05, we do not reject. Constant variance is reasonable.

\[
\text{Step two: individual } t\text{-tests with correction for multiple comparisons}
\]

- If we reject the overall F-test, we proceed to further analysis.

- The most common tests of interest are ‘all pairwise comparisons’ \( (\mu_1 \text{ vs. } \mu_2, \mu_1 \text{ vs. } \mu_3, \ldots, \mu_{m-1} \text{ vs. } \mu_m) \).

- We can use the Bonferroni correction to make sure the set of all \( \binom{m}{2} \) tests is done at an overall \( \alpha \) level error rate.

  - For this correction, we can multiply each regular pairwise \( t \)-test p-value by the number of tests to get the adjusted p-value.

  - Then, compare each adjusted p-value to \( \alpha \).

In R, we can use the ‘pairwise.t.test’ function to do all pairwise comparisons.

\[
\text{>> R Documentation: ‘pairwise.t.test’}
\]

\#
\text{Calculate pairwise comparisons between group levels}
\text{with corrections for multiple testing}

\[
\text{> pairwise.t.test(rate,group,p.adjust.method="bonferroni")}
\]

Pairwise comparisons using \( t \) tests with pooled SD

data: rate and group

\[
\begin{align*}
\text{C} & \quad \text{F} & \quad \text{P} \\
0.037 & \quad \text{0.031} & \quad \text{1.2e-05}
\end{align*}
\]

\[
P \text{ value adjustment method: bonferroni}
\]

All three p-values are compared against \( \alpha = 0.05 \), and in this case, ALL groups are significantly different from each other.

The probability that we have falsely rejected \( H_0 \) for 1 or more of these 3 tests is \( \leq 0.05 \).

Looking back at the individual \( t \)-tests comparing F and P to the baseline group C:

\[
\text{> summary(lm.out)}
\]

Coefficients:

\[
\begin{align*}
\text{(Intercept)} & \quad 82.524 & \quad 2.378 & \quad 34.709 & \quad <2e-16 \quad *** \\
\text{groupF} & \quad 8.801 & \quad 3.362 & \quad 2.617 & \quad 0.0123 \quad * \\
\text{groupP} & \quad -9.041 & \quad 3.362 & \quad -2.689 & \quad 0.0102 \quad *
\end{align*}
\]

We see that the adjusted p-values (below) are just 3 times the individual test-wise p-values above.

\[
\text{> pairwise.t.test(rate,group,p.adjust.method="bonferroni")}
\]

Pairwise comparisons using \( t \) tests with pooled SD

data: rate and group

\[
\begin{align*}
\text{C} & \quad \text{F} & \quad \text{P} \\
0.037 & \quad \text{0.031} & \quad \text{1.2e-05}
\end{align*}
\]

Plus, we get the comparison of Friends vs. Pet in the all-pairwise-comparisons statement.
You can change which group level is the baseline if this is useful to you.

Here’s the original, in alphabetical order:

```r
> contrasts(group)
  F P
C 0 0
F 1 0
P 0 1
```

```r
> levels(group)
[1] "C" "P" "P"
```

Here’s commands to switch Pet to the baseline:

```r
## Reassign baseline as the third level:
> contrasts(group)=contr.treatment(levels(group),base=3)
> contrasts(group)
  C F
C 1 0
F 0 1
P 0 0
```

There are also other multiple comparison adjustments available in `pairwise.t.test()`.

```r
> pairwise.t.test(rate,group,p.adjust.method="bonferroni")
```

Other options:

- holm
- hochberg
- hommel
- bonferroni
- BH
- BY
- fdr
- none