Chapter 4: Basic Analyses

Tyson S. Barrett

Summer 2017

Utah State University

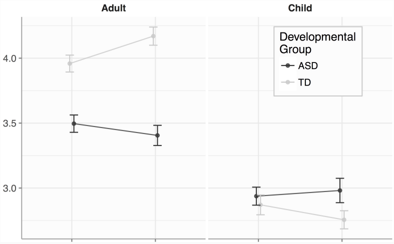
Introduction

Basic Analyses: The analyses taught in the first stats course

These include:

1. T-tests
2. ANOVA
3. Linear Regression

These allow us to assess relationships like that in the figure.



Maybe surprising:\ These all are doing essentially the same thing!

First, **T-TESTS!**

T-tests

Three Types

1. Simple
2. Independent Samples
3. Paired Samples

Three Types

Each will be demonstrated using:

```
df <- data.frame("A"=sample(c(0,1), 100, replace = TRUE),
                 "B"=rnorm(100),
                 "C"=rnorm(100))
df
```

```
    A           B           C
1   1  0.10374689 -0.32650412
2   0  1.56780670  0.77258383
3   1 -0.67845631  1.48447975
4   0  1.66587597 -1.29542982
5   1 -0.14670047  0.17361789
6   0 -0.62734606 -0.14871939
7   1  1.34815779 -1.39987953
8   1 -0.91928687 -0.27870608
9   0 -0.27793293 -0.38088161
```

Simple

Comparing a mean of a variable with μ.

```r
t.test(df$B, mu = 0)
```

```

	One Sample t-test

data:  df$B
t = 0.61766, df = 99, p-value = 0.5382
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.1403691  0.2672571
sample estimates:
 mean of x
0.06344402
```

Independent Samples

Comparing the means of two groups (dfA is the grouping variable).

```
t.test(df$B ~ df$A)
```

```

	Welch Two Sample t-test

data:  df$B by df$A
t = 0.40145, df = 93.178, p-value = 0.689
alternative hypothesis: true difference in means is not equal to
95 percent confidence interval:
 -0.3285636  0.4950772
sample estimates:
mean in group 0 mean in group 1 
     0.11006783      0.02681102 
```

Paired Samples

Comparing repeated measures (e.g., Pretest vs. Posttest).

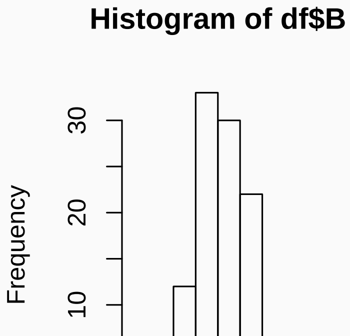
```r
t.test(df$B, df$C, paired = TRUE)
```

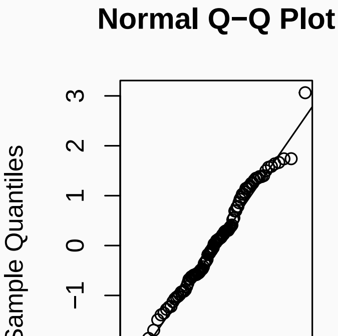
```

	Paired t-test

data:  df$B and df$C
t = 1.0152, df = 99, p-value = 0.3125
alternative hypothesis: true difference in means is not equal to
95 percent confidence interval:
 -0.1395149  0.4318630
sample estimates:
mean of the differences
               0.146174
```

Testing Assumptions of T-Tests

T-tests require that the data be normally distributed with approximately the same variance.

```
## Normality
par(mfrow = c(1,2))
hist(df$B)
qqnorm(df$B)
abline(a=0, b=1)
```





ANOVA

Analysis of Variance

The Analysis of Variance (ANOVA) is highly related to t-tests but can handle 2+ groups.

1. Provides the same p-value as t-tests
2. $t^2 = F$

For example:

```
fit_ano = aov(df$B ~ df$A)
summary(fit_ano)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
df$A         1   0.17  0.1708   0.161   0.69
Residuals   98 104.28  1.0641
```

```
t.test(df$B ~ df$A)$p.value
```

```
[1] 0.6890044
```

Analysis of Variance

```r
fit_ano = aov(df$B ~ df$A)
summary(fit_ano)
t.test(df$B ~ df$A)$p.value
```

Notice in the code:

- We assigned the `aov()` the name `fit_ano` (which we could have called anything)
- We used the `summary()` function to see the F and p values.
- We pulled the p-value right out of the `t.test()` function.

1. One-Way
2. Two-Way (Factorial)
3. Repeated Measures
4. A combination of Factorial and Repeated Measures

Types

We will use the following data set for the examples:

```
df <- data.frame("A"=sample(c(0,1), 100, replace = TRUE),
                 "B"=rnorm(100),
                 "C"=rnorm(100),
                 "D"=sample(c(1:4), 100, replace = TRUE))
df
```

```
  A            B            C D
1 0  1.815164862 -0.745933298 3
2 1  0.319942259 -0.761285771 4
3 0 -0.607921175 -0.141720206 4
4 0 -0.784290115 -0.205211852 4
5 0 -1.651797059  1.461840525 2
6 0  0.711887211 -0.003306458 1
7 0 -0.108334007  0.340000502 3
8 0  0.502882891  0.622497804 1
9 1 -1.732095717  0.122020327 1
```

One-Way

A One-Way ANOVA can be run using `aov()`.

```
fit1 = aov(B ~ D, data = df)
summary(fit1)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
D            1   0.08  0.0760   0.076  0.783
Residuals   98  97.40  0.9939
```

Two-Way

A Two-Way ANOVA uses essentially the exact same code with a minor change—including the other variable in an interaction.

```
fit2 = aov(B ~ D * A, data = df)
summary(fit2)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
D            1   0.08  0.0760   0.075  0.785
A            1   0.08  0.0825   0.081  0.776
D:A          1   0.10  0.1029   0.102  0.751
Residuals   96  97.21  1.0126
```

The `D:A` line highlights the interaction term whereas the others show the main effects.

Repeated Measures

Combination

Checking Assumptions

Linear Regression