1 Comparing Item Properties

Here we will look at how well our items are matched on Word Frequency and Bigram Frequency. The mean Word Frequency for HighN words is 69.44 and for LowN words the mean is 74.96. Here we use a t-test to make sure that those are not significantly different.

```r
> t.test(CELEX ~ G, Items)
Welch Two Sample t-test
data:  CELEX by G
t = -0.4979, df = 228.114, p-value = 0.619
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -27.39179   16.34113
sample estimates:
 mean in group hiN  mean in group loN
      69.43675       74.96208
```

The non-significant t-value tells us that the HiN and LoN items were well matched on Frequency. But just because two sets of items have similar means, it doesn’t necessarily mean that those two sets are well matched. It could be that the spread is much greater in one case than the other. This will result in similar means and, due to the large differences in variance, non-significant differences. Thus, it is important to visually inspect how similar/different the two sets of items are. Here we do that using density plots.
From that we can see that the two distributions are well matched, not only in their means, but throughout the whole range of values. We can use the Kolmogorov-Smirnov test to further test for differences in the two distributions.

```r
> ks.test(jitter(hiN$CELEX), jitter(loN$CELEX))

Two-sample Kolmogorov-Smirnov test

data: jitter(hiN$CELEX) and jitter(loN$CELEX)
D = 0.05, p-value = 0.9983
alternative hypothesis: two-sided
```

The lack of a significant difference here indicates that it is unlikely that these two distributions are different.

To demonstrate how two sets of items can have similar mean values, even while coming from two very different distributions, let’s replace a few of the Frequency values for the HighN words with the mean value for that set.

```r
> newhi <- sort(hiN$CELEX)
> newhi[10:30] <- rep(mean(newhi), 21)
```

Before we check to see if these new values are different from the old values and/or from the LowN items, let’s make sure that we actually changed the distribution. To do that, we plot these new values against the old values.
Yes, those are quite different. And the Kolmogorov-Smirnov test confirms this.

Two-sample Kolmogorov-Smirnov test

data: jitter(newhi) and jitter(loN$CELEX)
D = 0.225, p-value = 0.004599
alternative hypothesis: two-sided

OK, so this new distribution is different. But what about the means? The mean frequency value for this new set is 79.19, while the mean frequency value for the original set is 69.44. A t-test shows that these new frequency values are not different from the original set. Importantly for the present purposes, a second t-test shows that the new values are not different from the LowN items either.

Welch Two Sample t-test

data: hiN$CELEX and newhi
t = -1.0156, df = 237.194, p-value = 0.3109
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-28.659113 9.161751
sample estimates:
mean of x mean of y
69.43675 79.18543

Welch Two Sample t-test
data: newhi and loN$CELEX
t = 0.3892, df = 222.573, p-value = 0.6975
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  -17.16384  25.61053
sample estimates:
mean of x  mean of y
  79.18543  74.96208

So here is a situation where we have items coming from two very different distributions despite
sharing similar means.

The upshot? It is not enough to rely on a conventional t-test to tell you if your two sets of val-
ues come from the same distribution or not.

2 Types of t-tests

In the section above, we used a t-test to compare two vectors of equal length. But there are a
few different versions of the t-test and it is important to use the appropriate one. The first is
the one-sample t-test. This let’s you test whether your vector of values is different from zero (the
default) or from any other value you specify. Here is how to run a one-sample t-test:

> t.test(newhi, mu = 0)

    One Sample t-test

data: newhi
t = 12.0218, df = 119, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
  66.14286  92.22801
sample estimates:
mean of x
  79.18543

In the section above we ran a two-sample t-test. For that test, we used the default setting
(var.equal = FALSE) and did not assume that our two samples had equal variances. In that
default case, the Welch t-test is used, where the degrees of freedom are adjusted for the unequal
variances. When we assume equal variance (var.equal = TRUE), the student’s t-test is used. Here
we will compare the results of two t-tests. First we use the default setting, where we assume
unequal variances:

> t.test(CELEX ~ G, Items)

    Welch Two Sample t-test

data: CELEX by G
t = -0.4979, df = 228.114, p-value = 0.619
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  -27.39179  16.34113
sample estimates:
mean in group hiN  mean in group loN
  69.43675        74.96208
Now we will assume equal variances:

```r
> t.test(CELEX ~ G, var.equal = TRUE, Items)
```

Two Sample t-test

data: CELEX by G
t = -0.4979, df = 238, p-value = 0.619
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-27.38695 16.33628
sample estimates:
mean in group hiN mean in group loN
69.43675 74.96208

Note that in both of these, we are running a 'two-tailed' test. That is, we do not assume a
direction (e.g. x < y), and so we are simply testing the possibility that our t-value is more extreme
(in either tail of the t-distribution) than what would be expected under the null hypothesis. In
some cases, there is very strong independent evidence that your 'x' should be less than your 'y'
and, in this case, you can run a one-tailed test.

```r
> t.test(CELEX ~ G, alternative = "less", var.equal = TRUE, Items)
```

Two Sample t-test

data: CELEX by G
t = -0.4979, df = 238, p-value = 0.3095
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
-Inf 12.79955
sample estimates:
mean in group hiN mean in group loN
69.43675 74.96208

2.1 paired-samples t-test

In a 'repeated-measures' design, where you test each subject twice, once in ConditionA and once
in ConditionB, it is appropriate to test for the difference between conditions. This can be done in
two ways. Conceptually, the easiest is to take the differences between conditions (i.e. ConditionB
- ConditionA) and then run a one-sample t-test over the differences to see if the differences are
equal to zero (null hypothesis) or not. We will do that below:

```r
> x <- hiN$CELEX
> y <- loN$CELEX
> t.test(x - y, mu = 0)
```

One Sample t-test

data: x - y
t = -0.5079, df = 119, p-value = 0.6124
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
-27.06511 16.01444
sample estimates:
mean of x
-5.525333
Alternatively, you can specify `paired = TRUE` in the call to t.test:

```r
> t.test(CELEX ~ G, paired = T, Items)

Paired t-test
data: CELEX by G
t = -0.5079, df = 119, p-value = 0.6124
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  -27.06511 16.01444
sample estimates:
mean of the differences
  -5.525333
```

Perhaps the best way to understand the differences between paired- and independent-samples t-tests is to have a look at how the t-value is calculated in each case. Here is the way to calculate the t-value for independent samples:

```r
> independent.t.test <- function(x, y) {
+   return((mean(x) - mean(y))/sqrt(var(x)/length(x) + var(y)/length(y)))
+ }
```

And here is the function for calculating the t-value for paired samples:

```r
> paired.t.test <- function(x, y) {
+   d = x - y
+   n = length(x)
+   return(sum(d)/sqrt((n * (sum(d^2)) - sum(d)^2)/(n - 1)))
+ }
```

Finally, we show how linear regression and anova return the same result as an independent-samples t-test that assumes equal variances:

```r
> summary(lm(CELEX ~ G, Items))

Call:
  lm(formula = CELEX ~ G, data = Items)

Residuals:
   Min     1Q   Median     3Q    Max
-66.86  -54.57  -37.45   12.36  381.52

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  69.4370     7.8475   8.849  <2e-16 ***
GloN         5.5253     11.0970   0.498  0.619
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 85.96 on 238 degrees of freedom
Multiple R-squared: 0.001041,   Adjusted R-squared: -0.003157
F-statistic: 0.2479 on 1 and 238 DF,  p-value: 0.619
```

```r
> summary(aov(CELEX ~ G, Items))
```
Df  Sum Sq Mean Sq  F value  Pr(>F)
G       1  1832  1831.8   0.2479  0.619
Residuals 238 1758606  7389.1

> t.test(CELEX ~ G, var.equal = TRUE, Items)

Two Sample t-test

data: CELEX by G
t = -0.4979, df = 238, p-value = 0.619
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-27.38695   16.33628
sample estimates:
mean in group hiN mean in group loN
69.43675        74.96208