C-N Backpack Analysis
11-2021

```
dads@Fergerson: ~
File Edit View Search Terminal Help
 [26] 30 10 10 10 30 10 20 20 30 10 20 10  5 20 10 10 10 20 10 60 10 20 20 10 10
 [51] 10 30 20 20 60 10 30 20 10 50 30 20 10  5 20 10 10 30 60 30 10 30 60 10 10
 [76] 20 30 20 30 10 20 10 40 10 30  5 30 20 30 30  5 10 10 20 30 10 20 10 10 10
[101] 20 10 10 10 60 60 10 10 30 10 20 30 10 10 20 10 10 10 20 20 10  5 20 10 10
> proper = length(which(wearpr == 20));
> proper;
[1] 33
> prop.test(33,125, p=.5, alternative = "less", conf.level = .95);

        1-sample proportions test with continuity correction

data:  33 out of 125, null probability 0.5
X-squared = 26.912, df = 1, p-value = 1.065e-07
alternative hypothesis: true p is less than 0.5
95 percent confidence interval:
 0.0000000 0.3375704
sample estimates:
    p
0.264

> wtproper = length(which(cnbp$WTFR == 5));
> wtproper;
[1] 95
> prop.test(95,125, p=.5, alternative = "less", conf.level = .95);

        1-sample proportions test with continuity correction

data:  95 out of 125, null probability 0.5
X-squared = 32.768, df = 1, p-value = 1
alternative hypothesis: true p is less than 0.5
95 percent confidence interval:
 0.0000000 0.8204647
sample estimates:
   p
0.76

>
```

RQ 1
We are interested in the ratio of students who carry the backpack in a totally correct way. This entails ensuring that the backpack wt is no more than 10% of the body wt, worn with two straps and has an appropriate back position. In order to assess this, we need a variable product … specifically we want to see student observations for which the WTFR = 5 (proportion of wt < 10%), STR = 2 and WORN = 2 (correct position = yes). The product of these numbers is 20. From the R output above, we see that 33 of the 125 observations resulted in a value of 20.

1. H: $\pi = .5$, K: $\pi < .5$, $\alpha = .05$
2. Conditions met: $n\pi = n(1-\pi) = 62.5 > 10$, population ≈ 1800 > 10n = 1250.
   $X^2 \sim \chi^2(1)$
3. $X^2 = 26.912$
4. $P \approx .0000001$
5. Reject H. Based on the data we can conclude that the proportion of C-N students correctly wearing backpacks (of those wearing backpacks) is less than half.

Conditions for CI met: successes = 33 > 10, failures = 92 > 10, population ≈ 1800 > 10n = 1250.
$CI_{.95} = (0, .34)$

RQ 2

We are interested in the ratio of students who carry the backpack with backpack wt no more than 10% of the body wt. This is effectively a subset question of RQ1. We are seeking student observations in which the WTFR = 5 (proportion of wt < 10%). From the R output above, we see that 95 of the 125 observations resulted in a WTFR value of 5.

1. H: $\pi = .5$, K: $\pi < .5$, $\alpha = .05$
2. Conditions met: $n\pi = n(1-\pi) = 62.5 > 10$, population $\approx 1800 > 10n = 1250$.
   $X^2 \sim \chi^2\,(1)$
3. $X^2 = 32.768$
4. $P \approx 1$
5. Fail to reject H. Based on the data we cannot conclude that the proportion of C-N students wearing appropriately weighted backpacks (of those wearing backpacks) is less than half. In fact, the true proportion is quite likely well over half (p = .76).

Conditions for CI met: successes = 95 > 10, failures = 30 > 10, population ≈ 1800 > 10n = 1250.
$CI_{.95} = (0, .82)$

```
dads@Fergerson: ~
File  Edit  View  Search  Terminal  Help
119    3        2   0   2.5      1   9.80    5   2    2
120    9        2   B   7.0      1  16.66    5   2    2
121    3        2   T   5.0      1  11.04    5   2    1
122   10        2   T   3.5      1  11.72    5   1    1
123    3        1   T   4.0      1   8.58    5   2    2
124    4        1   T   4.0      1  10.06    5   2    1
125    2        1   T   6.0      1   9.58    5   2    1
> bppain = length(which(cnbp$AFTER == 2));
> bppain;
[1] 48
> prop.test(48,125, p=.5, alternative = "greater", conf.level = .95);

        1-sample proportions test with continuity correction

data:  48 out of 125, null probability 0.5
X-squared = 6.272, df = 1, p-value = 0.9939
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval:
 0.3118469 1.0000000
sample estimates:
    p
0.384

>
```

RQ3

Here, we are interested in the proportion of students reporting back pain after carrying the backpack. We are seeking student observations in which the variable AFTER = 2 (PPP - post pack pain). From the R output above, we see that only 48 of the 125 observations resulted in an AFTER value of 2.

RQ3 (cont)

1. H: $\pi = .5$, K: $\pi > .5$, $\alpha = .05$
2. Conditions met: $n\pi = n(1-\pi) = 62.5 > 10$, population $\approx 1800 > 10n = 1250$.
   $X^2 \sim \chi^2\ (1)$
3. $X^2 = 6.272$
4. $P \approx .9939$
5. Fail to reject H. Based on the data we cannot conclude that the proportion of C-N students wearing experiencing pain after wearing backpacks (of those wearing backpacks) is more than half. In fact, the true proportion is quite likely well under half ($p = .384$).

Conditions for CI met: successes = 48 > 10, failures = 77 > 10, population $\approx 1800 > 10n = 1250$.
$CI_{.95} = (.31, 1)$

RQ4

Here we are concerned with relationships, and as such, we wish to find those that exist between various demographic factors and the proper weight and/or carriage of the backpack. We'll begin by looking at back pain (AFTER) as it relates to backpack wt ratio (WTFR). In the output below, the matrix necessary for the Chi Square test for independence has the problem of too few observations in the last two categories for those who have pain. So we'll lump those observations into the second category making a 2 x 2 matrix (which will meet the Chi Square distribution conditions).

```
dads@Fergerson: ~
File Edit View Search Terminal Help
>
>
> afterwtr <- table(cnbp$AFTER,cnbp$WTFR);
> afterwtr;

     5 15 20 25
  1 67 10  0  0
  2 28 18  1  1
> afterwtr = matrix(c(67,10,28,20), nrow = 2, ncol = 2, byrow = T);
> chisq.test(afterwtr);

        Pearson's Chi-squared test with Yates' continuity correction

data:  afterwtr
X-squared = 11.8075, df = 1, p-value = 0.0005899

> chisq.test(afterwtr) $expected;
      [,1]  [,2]
[1,] 58.52 18.48
[2,] 36.48 11.52
>
```

Now the test can be performed.

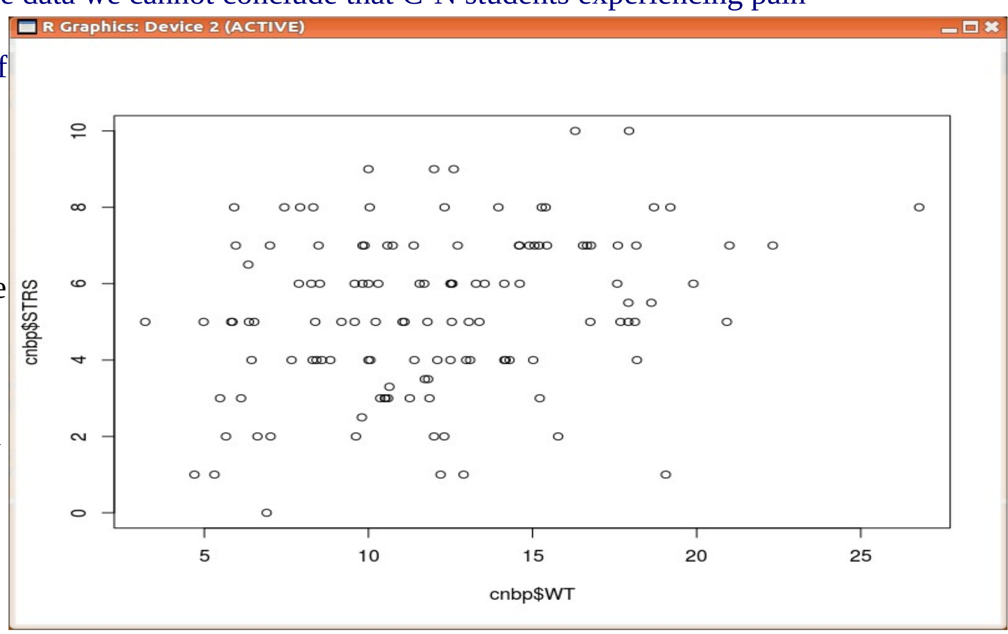
1. H: variables AFTER and WTFR are independent, K: not H, $\alpha = .05$
2. Conditions met: All expected matrix values ≥ 5, sampling is independent.
   $X^2 \sim \chi^2\ (1)$
3. $X^2 = 11.8075$
4. $P \approx .0006$
5. Reject H. Based on the data we conclude that C-N students experiencing pain after wearing backpacks does appear to be related to excessive backpack weight.

RQ5
We'll continue by looking at back pain (AFTER) as it relates to correct carriage of the backpack (wearpr). In the output below, the matrix necessary for the Chi Square test for independence has the problem of too few observations in the last two categories for those who have pain. Once again we'll lump those observations into a 2 x 2 matrix (basically improper vs proper wear; this will meet the Chi Square distribution conditions).

```
dads@Fergerson: ~
File  Edit  View  Search  Terminal  Help
[101] 20 10 10 10 60 60 10 10 30 10 20 30 10 10 20 10 10 10 20 20 10  5 20 10 10
> afterwr <- table(cnbp$AFTER,wearpr);
> afterwr;
   wearpr
     5 10 20 30 40 50 60
  1  5 37 25  8  0  0  2
  2  1 19  8 14  1  1  4
> afterwr = matrix(c(52,25,40,8), nrow = 2, ncol = 2, byrow = T);
> afterwr;
     [,1] [,2]
[1,]   52   25
[2,]   40    8
> chisq.test(afterwr);

        Pearson's Chi-squared test with Yates' continuity correction

data:  afterwr
X-squared = 3.0296, df = 1, p-value = 0.08176

> chisq.test(afterwr) $expected;
       [,1]   [,2]
[1,] 56.672 20.328
[2,] 35.328 12.672
>
```

Now the test can be performed.

1. H: variables AFTER and wearpr are independent, K: not H, $\alpha = .05$
2. Conditions met: All expected matrix values ≥ 5, sampling is independent.
   $X^2 \sim \chi^2(1)$
3. $X^2 = 3.03$
4. $P \approx .08$
5. Fail to Reject H. Based on the data we cannot conclude that C-N students experiencing pain after wearing backpacks does not appear to be related to proper wear of the backpack (although there is mildly significant evidence to the contrary).



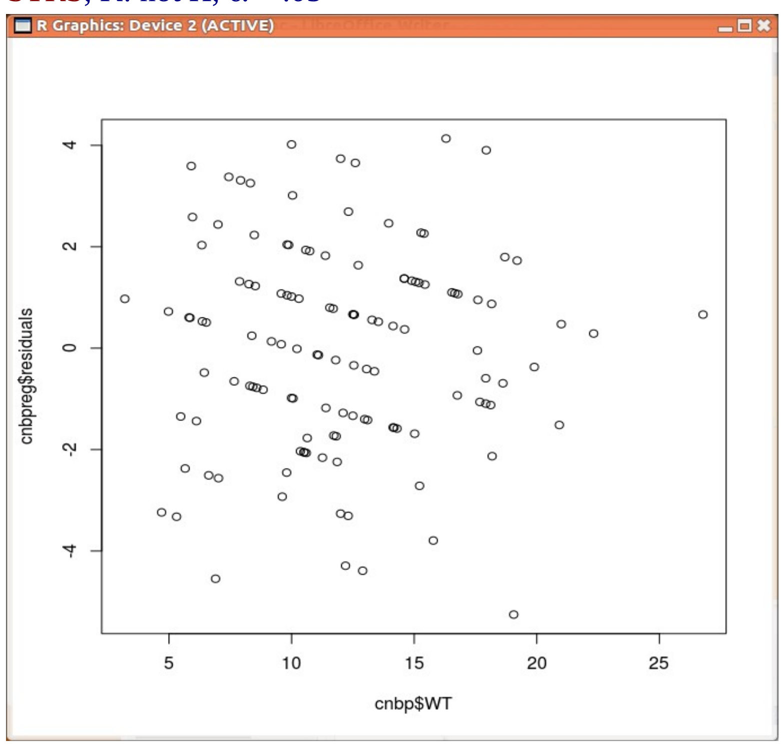
RQ6
For question 6 we want to investigate the relationship student stress (STRS) and backpack weight (WT). In this case, we perform a simple linear regression for the test and then check the conditions for the F distribution afterward. Here is the intial plot of both variables from R.

We suspect that STRS is a function of WT in this case, hence the order of the plot and the test.

```
dads@Fergerson: ~
File  Edit  View  Search  Terminal  Help
> anova (lm(cnbp$STRS ~ cnbp$WT));
Analysis of Variance Table

Response: cnbp$STRS
           Df Sum Sq Mean Sq F value   Pr(>F)
cnbp$WT     1  45.43  45.432  11.264 0.001051 **
Residuals 123 496.09   4.033
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> summary (lm(cnbp$STRS ~ cnbp$WT));

Call:
lm(formula = cnbp$STRS ~ cnbp$WT)

Residuals:
    Min      1Q  Median      3Q     Max
-5.2538 -1.4172  0.2454  1.2880  4.1336

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.57823    0.53410   6.700 6.68e-10 ***
cnbp$WT      0.14038    0.04183   3.356  0.00105 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.008 on 123 degrees of freedom
Multiple R-squared: 0.0839,     Adjusted R-squared: 0.07645
F-statistic: 11.26 on 1 and 123 DF,  p-value: 0.001051
```

Now the test can be performed.

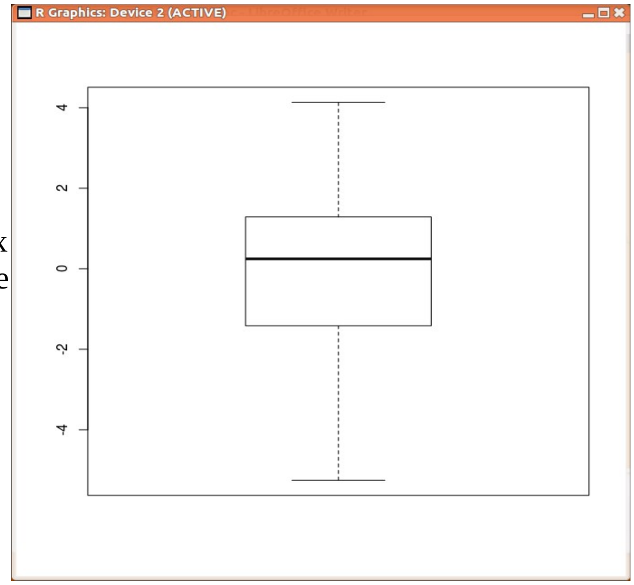
1. H: No linear relationship between WT and STRS, K: not H, α = .05
2. Conditions met: See below.
   W ~ F (1, 123)
3. W = 11.264
4. P ≈ .001
5. Reject H. Based on the data we can conclude that C-N students very likely experience stress in linear relation to the weights of their backpack. Our guess is that the weight is related to the number of hours being carried … which is really the root cause of the stress.



A view of the residuals in conjunction with the “independent” variable reveals a nice rectangular shape for meeting the homogeneity of variance assumptions for an F distribution. The ensuing boxplot shows a nice normal distribution of the same residuals indicating that we have met the normality conditions for an F distribution in step 2 of the hypothesis test.

RQ7

We'll continue by looking at backpack weight (WT) and its relationship to the carrying of ebooks (EB) by performing a two sample t test. In the output below, the matrix necessary for the Chi Square test for independence has the problem of too few observations in the last two categories for those who have pain. Once again we'll lump those observations into a 2 x 2 matrix (basically improper vs proper wear; this will meet the Chi Square distribution conditions).



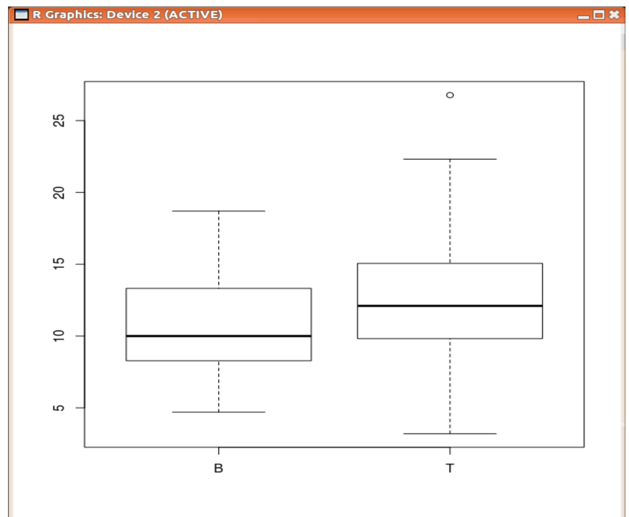
Now the test can be performed.

1. H: backpack wt will not differ by the use of ebooks (EB), K: ebook usage will result in lower backpack wt (WT), $\alpha = .05$
2. Conditions met: Independent samples, both samples with slightly right skewed distributions (see parallel boxplots at right), both sample sizes > 30.

   $t = (X_{ebmean} - X_{tbmean})/SE \sim t(55.023)$
3. $t = -1.68$
4. $P \approx .0498$
5. Reject H. Based on the data we can conclude that C-N students using ebooks likely have lower backpack weight.

Conditions for CI met: see step 2 above.

$CI_{.95}$ = (-infi, -.002)lb



dads@Fergerson: ~

File Edit View Search Terminal Help

```
122  10       2  T  3.5     1 11.72    5   1    1
123   3       1  T  4.0     1  8.58    5   2    2
124   4       1  T  4.0     1 10.06    5   2    1
125   2       1  T  6.0     1  9.58    5   2    1
> boxplot(cnbp$WT ~ cnbp$EB);
> t.test(cnbp$WT ~ cnbp$EB, alternative = "less", conf.level = .95);

        Welch Two Sample t-test

data:  cnbp$WT by cnbp$EB
t = -1.6751, df = 55.023, p-value = 0.04979
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
         -Inf -0.001789248
sample estimates:
mean in group B mean in group T
       10.95226        12.38011

>
```