Moodle Questions Chapter 10

Homer White

Thursday, June 25, 2014

Bee LifeSpans

Recall the data frame `BeeLifespans` from the `abd` package:

```
require(abd)
data(BeeLifeSpans)
View(BeeLifespans)
help(BeeLifespans)
```

The variable **hours** records the life-span, in hours, of each of 33 foraging honey bees. We may think of them as a random sample from the population of all foraging honey bees.

We know of another species of bees than lives 19 hours, on average, and we wonder if foraging honey bees differ, in their mean lifespan, from this other species. Hence we perform a two-sided test of significance.

We first define:

mu = the mean lifespan, in hours, of all foraging honey bees.

Then we set up our hypotheses:

H0: mu = 19

Ha: mu != 19 (the "!= means"not equal to")")

Then we run our test:

```
ttestGC(~hours,data=BeeLifespans, mu=19)
```

```
##
##
## Inferential Procedures for One Mean mu:
##
##
## Descriptive Results:
##
##   variable   mean     sd   n
##      hours  27.85  20.56  33
##
##
## Inferential Results:
##
## Estimate of mu:    27.85
## SE(x.bar):    3.578
##
## 95% Confidence Interval for mu:
##
##          lower.bound          upper.bound
##          20.559809            35.137161
##
```

```
## Test of Significance:
##
##  H_0:  mu = 19
##  H_a:  mu != 19
##
##  Test Statistic:     t = 2.473
##  Degrees of Freedom:  32
##  P-value:        P = 0.0189
```

Which of the following statements are correct, as part of the commentary you would provide for this study?

We conclude that the data provides strong evidence that the mean lifespan for all forging honey bees is more than 19 hours.

Since the P-value is less than 0.05, we reject the Null Hypothesis.

The Null Hypothesis expects that the sample mean should be around 19, give or take some for chance error, but the sample mean lifespan in this study was 3.578 standard errors above what the Null Hypotheses expected.

Feedback: No, it was 2.473 SE's above what the Null expects. The 3.578 is the Standard error itself.

The Null's belief (19) lies entirely below the 95%-confidence interval. Based on the data, it's not reasonable to believe that mu is 19.

We conclude that it is absolutely certain that the mean lifespan for all foraging honey bees is more than 19 hours.

The P-value is 0.0189, so there is about a 1.89% chance that the Null Hypothesis is correct.

Feedback: No, the Null is either right or wrong: no "chances" about it. The P-value tells you that if the Null were right, then there would be about a 1.89% chance of getting a test statistic at least as far form 0 as the one we actually got.

Beans: Parameter and Hypotheses

In the `tigerstats` package we find the `beans` data frame:

```
data(beans)
View(beans)
help(beans)
```

This was a repeated measures-experiment performed at UC-Davis; fifteen students participated. Each student was asked to place as many beans into a cup as he/she could, in 15 seconds. Each student performed this task once with the dominant hand, and once with the non-dominant hand, but the order of performance was randomized. Assume that the subjects in the experiment are the fifteen students. (*Reminder about terminology*: your dominant hand is the hand you use the most.)

The Research Question was:

> *Which hand is more dextrous: the dominant hand, or the non-dominant hand?*

We would like to perform a *two-sided test* of significance to answer this question. Which of the following options is the best way to define the parameter of interest and to state Null and Alternative hypotheses? (Note: "!=" stands for "not equal to")

Option A:

Let mu-d = the mean difference in the number of means moved (dominant hand minus non-dominant hand), for all students at UC-Davis.

H0: mu-d = 0 Ha: mu_d != 0

Option B:

Let mu-d = the mean difference in the number of means moved (dominant hand minus non-dominant hand), for all students at UC-Davis.

H0: mu-d != 0 Ha: mu_d = 0

Feedback: This mixes up the Null and Alternative. The Null is always the hypotheses that states an equality.

Option c:

H0: mu-d = 0 Ha: mu_d != 0

Feedback: Nobody can tell what the hypotheses mean unless you first define the parameter mu-d.

Option D:

Let mu-d = the mean difference in the number of means moved (dominant hand minus non-dominant hand), for all students at UC-Davis.

H0: mu-d = 0 Ha: mu_d > 0

Feedback: This is a one-sided test! (Directions asked for two-sided.)

Beans: Thinking about Safety

Can we rely on the results of a t-test for the `beans` study? There were fifteen students serving as subjects, and let's say that the researchers chose them randomly from the population.

Choose the best option below.

Option A:

The sample size is not very large: we had better make a histogram (or box-plot) of the number of beans moved by the dominant hand, and check it for skewness and outliers.

Feedback: You are working with the differences (that's the variable **Diff** in the data frame). You want a histogram or box-plot of **Diff**.

Option B:

The sample size is not very large: we had better make a histogram (or box-plot) of the number of beans moved by the non-dominant hand, and check it for skewness and outliers.

Feedback: You are working with the differences (that's the variable **Diff** in the data frame). You want a histogram or box-plot of **Diff**.

Option c:

Since there are fifteen students, each measured twice, there are thirty total measurements. That's a big-enough sample size, so we are probably safe to proceed without looking at graphs of the data.

Feedback: You are working with the differences (that's the variable **Diff** in the data frame). There is one difference for each of the fifteen students, so the sample size here is fifteen, not thirty.

Option D:

The sample size is not very large: we had better make a histogram (or box-plot) of the differences in the number of beans moved by the two hands, and check the plot for skewness and outliers.

Feedback: Good call: the sample size was 15, below our usual "safety" cut-off of 30.

Beans: The R-code

Thinking about the `beans` study: which of the following bits of R-code correctly performs a two-sided test of significance to answer our Research Question?

Option A:

```
ttestGC(~Dom - NonDom, data=beans, mu=0)
```

Option B:

```
ttestGC(~Dom - NonDom, data=beans, mu=0,
        alternative="less")
```

Feedback: No, that's a one-sided "less-than" test.

Option C:

```
ttestGC(Dom - NonDom, data=beans, mu=0)
```

Feedback: R will throw an error at you, because you are missing the ~ sign that indicates your formula in formula-data input.

Option D:

```
ttestGC(~NonDom - Dom, data=beans, mu=0)
```

Feedback: Close, but it's the wrong way around, considering that we defined mu-d as the mean of differences (dominant minus non-dominant hand).

Option E:

```
ttestGC(~dom - nondom, data=beans, mu=0)
```

Feedback: R will throw an error: the variables "dom" and "nondom" are misspelled. Remember that R is case-sensitive!

Beans: How many SE's?

Still working with the `beans` study. The output for the two-sided test of significance is as follows:

```
##
##
## Inferential Procedures for the Difference of Means mu-d:
##   NonDom minus Dom
##
##
## Descriptive Results:
##
##    Difference mean.difference sd.difference  n
##  Dom - NonDom          1.067         2.434 15
##
```

```
##
## Inferential Results:
##
## Estimate of mu-d:      1.067
## SE(d.bar):    0.6284
##
## 95% Confidence Interval for mu-d:
##
##          lower.bound         upper.bound
##          -0.281175           2.414509
##
## Test of Significance:
##
##  H_0:  mu-d = 0
##  H_a:  mu-d != 0
##
##  Test Statistic:      t = 1.697
##  Degrees of Freedom:   14
##  P-value:         P = 0.1117
```

Looking at the output, which of the following remarks is correct?

Option A:

A believer in the Null Hypotheses would expect d-bar to be about 0, give or take some for chance error. d-bar was actually 1.067, which is about 1.7 standard errors below what the Null expects. This result is a bit surprising if the Null is true, but it's not all that unlikely to occur.

Option B:

A believer in the Null Hypotheses would expect d-bar to be about 0, give or take some for chance error. d-bar was actually 1.067, which is about 1.7 standard errors below what the Null expects. This result is very unlikely to occur, if the Null is true.

Feedback: Results within two SE's of what's expected aren't generally considered very surprising, and usually do not have a very small chance of happening.

Option c:

A believer in the Null Hypotheses would expect d-bar to be about 0, give or take some for chance error. d-bar was actually 1.067, which is about 0.6284 standard errors below what the Null expects. This result not at all unlikely to occur, if the Null is true.

Feedback: 0.6284 is the standard error itself. The t-statistic is what tells you how many SE's the d-bar is above or below what the Null expects.

Beans: Interpreting the P-value

Still working with the beans study. The output for the two-sided test of significance is as follows:

```
##
##
## Inferential Procedures for the Difference of Means mu-d:
##   NonDom minus Dom
##
##
```

```
## Descriptive Results:
##
##    Difference mean.difference sd.difference  n
##  Dom - NonDom               1.067          2.434 15
##
##
## Inferential Results:
##
## Estimate of mu-d:     1.067
## SE(d.bar):    0.6284
##
## 95% Confidence Interval for mu-d:
##
##          lower.bound           upper.bound
##          -0.281175             2.414509
##
## Test of Significance:
##
##  H_0:  mu-d = 0
##  H_a:  mu-d != 0
##
##  Test Statistic:     t = 1.697
##  Degrees of Freedom:   14
##  P-value:      P = 0.1117
```

Which of the following statements below is correct, as a way of explaining what the P-value means?

Option A:

If the hand you use makes no difference in how many beans you can move, then there is about an 11.17% chance of getting a test-statistic at least as far from 0 as the 1.697 (the statistic we actually got).

Option B:

There is about an 11.17% chance that the Null Hypothesis is correct.

Option C:

About 11.17% of the differences were bigger than what the null Hypothesis would expect.

Option D:

If you can move more beans with your dominant hand, then there is about an 11.17% chance of getting a test-statistic at least as far from 0 as the 1.697 (the statistic we actually got).

Beans: Decision and Conclusion

Recall again the output for the beans study.

```
##
##
## Inferential Procedures for the Difference of Means mu-d:
##   NonDom minus Dom
##
##
## Descriptive Results:
```

```
##
##    Difference mean.difference sd.difference  n
##  Dom - NonDom            1.067          2.434 15
##
##
## Inferential Results:
##
## Estimate of mu-d:      1.067
## SE(d.bar):    0.6284
##
## 95% Confidence Interval for mu-d:
##
##          lower.bound         upper.bound
##          -0.281175            2.414509
##
## Test of Significance:
##
##  H_0:  mu-d = 0
##  H_a:  mu-d != 0
##
##  Test Statistic:     t = 1.697
##  Degrees of Freedom:   14
##  P-value:       P = 0.1117
```

Which of the following is correct as a decision-and-conclusion?

Option A:

Since the P-value is less than 0.05, we will reject the Null Hypothesis. This study provided strong evidence that you can move more beans with your dominant hand.

Feedback: The P-value was 0.1117, which is bigger than 0.05.

Option B:

Since the P-value is greater than 0.05, we will reject the Null Hypothesis. This study provided strong evidence that you can move more beans with your dominant hand.

Feedback: You reject when the P-value is smaller than your cut-off, not when it is bigger.

Option C:

Since the P-value is greater than 0.05, we will not reject the Null Hypothesis. This study provided strong evidence that it makes no difference which hand you use.

Feedback: Tests like this never give evidence FOR the Null hypothesis, even when the P-value is large.

Option D:

Since the P-value is greater than 0.05, we will not reject the Null Hypothesis. This study did not provide strong evidence that you can move more beans with your dominant hand.

Feedback: Good job. When the P-value is low, you reject the Null. The lower the P-value, the stronger your evidence against the Null and for the Alternative.

Napkins: R-code

Recall the `napkins` data frame from previous quiz questions:

```
data(napkins)
View(napkins)
help(napkins)
```

This was an observational study performed at Georgetown College. A small team of student researchers visited the Cafe several times during the lunch hour and recorded the sex of each student in the study as well as how many napkins each student used during the meal.

The Research Question was:

Who uses more napkins, on average, during lunch: a GC male or a GC female?

Let's say that we have defined the parameters as follows:

Let mu-1 = the mean number of napkins used at lunch by all GC males

Let mu-2 = the mean number of napkins used at lunch by all GC females.

Assume that we have some prior studies that suggest guys are messier eaters than gals are, so we would like to perform the following one-sided test:

H0: mu-1 - mu-2 = 0 (GC males and GC females use the same number of napkins, on average)

Ha: mu -1 - mu-2 > 0 (guys use more on average)

Which bit of R-code will do the job for us?

Option A:

```
ttestGC(napkins~sex,data=napkins,mu=0,
        alternative="greater",
        first="male")
```

Option B:

```
ttestGC(napkins~sex,data=napkins,mu=0,
        alternative="greater")
```

Feedback: Not quite. The groups will be in the wrong wrong order. As we defined the parameters, the males have to come first.

Option C:

```
ttestGC(napkins~sex,data=napkins,mu=0)
```

Feedback: This will perform a two-sided test. Also, the groups will be in the wrong wrong order. As we defined the parameters, the males have to come first.

Option D:

```
ttestGC(~ napkins - sex,data=napkins,mu=0,
        alternative="greater",
        first="male")
```

Feedback: R will throw an error. It does not make sense to try to study the difference between sex and the number of napkins used. Sex is not a number, and cannot be subtracted from the numerical variable **napkins**.

Testing Know-How

Which of the following are good things to keep in mind when you are performing a test of significance?

The smaller the P-value is, the stronger the evidence you have against the Null Hypothesis (and for the Alternative Hypothesis).

The bigger the P-value is, the weaker the evidence you have against the Null Hypothesis (and for the Alternative Hypothesis).

The bigger the P-value is, the stronger the evidence you have for the Null Hypothesis (and against the Alternative Hypothesis).

As a rule, we reject the Null Hypothesis when our P-value is less than 0.05.

As a rule, we reject the Null Hypothesis when our P-value is greater than 0.05.

Spider Speeds: Tests and Intervals

Remember the data frame `SpiderSpeed` from package `abd`?

```
require(abd)
data(SpiderSpeed)
View(SpiderSpeed)
help(SpiderSpeed)
```

Recall that in this study, the researchers took 32 *Tidarren* spiders and recorded how fast each one could run, in centimeters per second. Then they amputated a pedipalp from each one of the spiders, and again recorded how fast each spider could run. This was a *repeated measures* study!

The researchers wanted to know if the spiders could run faster, on average, without the hindrance of a pedipalp.

They decided to investigate their question by using their sample of spiders to make a 90% confidence interval for mu-d, the mean difference in speed (after amputation minus before amputation) for all *Tidarren* spiders.

They ran the following R-code and got their results:

```
ttestGC(~speed.after - speed.before, data=SpiderSpeed,
        conf.level = 0.90)
```

```
##
##
## Inferential Procedures for the Difference of Means mu-d:
##   speed.before minus speed.after
##
##
## Descriptive Results:
##
##                   Difference mean.difference sd.difference  n
##   speed.after - speed.before           1.186         1.074 16
##
##
## Inferential Results:
##
```

9

```
## Estimate of mu-d:     1.186
## SE(d.bar):    0.2684
##
## 90% Confidence Interval for mu-d:
##
##          lower.bound          upper.bound
##          0.715028             1.656222
```

Hmm, looks like we are 90%-confident that the mean post-amputation increase in speed for all *Tidarren* spiders is somewhere between 0.715 and 1.656 cm/sec.

Which of the following is correct?

Option A:

If they had performed a two-sided test of significance where the Null Hypothesis says that mu-d is 0.5, then the P-value would have been less than 0.10.

Option B:

If they had performed a two-sided test of significance where the Null Hypothesis says that mu-d is 1.3, then the P-value would have been more than 0.10.

Option C:

If they had performed a two-sided test of significance where the Null Hypothesis says that mu-d is 0.5, then the P-value would have been more than 0.10.

Option D:

If they had performed a two-sided test of significance where the Null Hypothesis says that mu-d is 1.3, then the P-value would have been less than 0.10.

Intervals from Tests

Researchers are interested in the mean mu of a certain population. They perform the following two-sided test:

H0: mu = 30

Ha: mu != 30

They get a P-value of 0.021.

Which of the following is true?

Option A:

If they had made a 95%-confidence interval for mu, then the interval would not have contained 30.

Option B:

If they had made a 99%-confidence interval for mu, then the interval would have contained 30.

Option C:

If they had made a 90%-confidence interval for mu, then the interval would not have contained 30.

Option D:

If they had made a 95%-confidence interval for mu, then the interval would have contained 30.

Option E:

If they had made a 99%-confidence interval for mu, then the interval would not have contained 30.

Option F:

If they had made a 90%-confidence interval for mu, then the interval would have contained 30.

Evidence for a Difference vs. Size of Difference

Suppose that is known that the mean time for an Alpaca to complete a 50-meter dash is 7.5 seconds. Let's also say that down in Bolivia, where Alpacas are bred for their excellent hair, people like to race Alpacas. There is a certain interest in vitamin supplements that might make then run faster.

A new supplement is developed and tested on 50,000 randomly-selected Alpacas. After taking the supplement for a period of six months, each Alpaca runs a 50-meter dash and its time is recorded.

The mean time turns out to be 7.49 seconds, and the SD of the times is 0.6 seconds. Researchers decide to perform a two-sided test of significance as follows:

Define mu = the mean 50-meter dash speed for all Alpacas, if all of them could have taken the supplement.

H0: mu = 7.5 (i.e., supplement makes no difference, on average) Ha: mu != 7.5 (i.e., supplement makes a difference, one way or the other)

Fro good measure, they construct a 95%-confidence interval for mu.

Here is the R-code for the test, along with the results:

```
ttestGC(mean=7.49,sd=0.6,
        n=50000,mu=7.5)
```

```
##
##
## Inferential Procedures for One Mean mu:
##
##
## Descriptive Results:
##
##   mean  sd      n
##   7.49 0.6 50000
##
##
## Inferential Results:
##
## Estimate of mu:    7.49
## SE(x.bar):     0.002683
##
## 95% Confidence Interval for mu:
##
##           lower.bound          upper.bound
##           7.484741             7.495259
##
## Test of Significance:
##
##   H_0:  mu = 7.5
##   H_a:  mu != 7.5
##
##   Test Statistic:     t = -3.727
##   Degrees of Freedom:  49999
```

```
##  P-value:         P = 0.0001942
```

Which of the following conclusions make the most sense?

Option A:

All values in the confidence interval are quite close to 7.5. What we have here is a tremendous amount of evidence (due to the large sample size) for a teeny-weeny improvement in speed from taking the supplement. It's probably not worth buying the supplement for your racing Alpaca, unless the supplement is very, very cheap or you expect your Alpaca to be in some mighty close races.

Option B:

Wow, check out that very low P-value of 0.001942! We've got extremely strong evidence here that taking the supplement increases speed, on average. Owners of racing Alpacas should feed this supplement to their Alpacas on a regular basis!

Types of Error: A

You take a sample of size 36 from a population, and compute the sample mean, which turns out to be 31.5. The standard deviation of the sample is 3. You want to test whether or not the mean mu of the population is 30, so you run this R-code:

```
ttestGC(mean=31.5,sd=3,n=36,mu=30)
```

Your plan is to reject the Null Hypothesis if the P-value is less than 0.05.

Unknown to you, the population mean mu is 30.9.

Which of the following is correct?

You made a Type-I Error.

You made a Type-II Error.

There was no error.

Types of Error: B

You take a sample of size 36 from a population, and compute the sample mean, which turns out to be 31.5. The standard deviation of the sample is 3. You want to test whether or not the mean mu of the population is 30, so you run this R-code:

```
ttestGC(mean=31.5,sd=3,n=36,mu=30)
```

Your plan is to reject the Null Hypothesis if the P-value is less than 0.05.

Unknown to you, the population mean mu is 30.

Which of the following is correct?

You made a Type-I Error.

You made a Type-II Error.

There was no error.

Types of Error: C

You take a sample of size 36 from a population, and compute the sample mean, which turns out to be 30.3. The standard deviation of the sample is 3. You want to test whether or not the mean mu of the population is 30, so you run this R-code:

```
ttestGC(mean=30.3,sd=3,n=36,mu=30)
```

Your plan is to reject the Null Hypothesis if the P-value is less than 0.05.

Unknown to you, the population mean mu is 30.1

Which of the following is correct?

You made a Type-I Error.

You made a Type-II Error.

There was no error.

Types of Error: D

In a certain population, the mean mu is 42. There are one thousand statisticians, and none of them know what mu is. Each statistician takes a simple random sample from this population, and each one uses his/her sample to perform the following two-sided test:

H0: mu = 42

Ha: mu != 42

How many of them will commit a Type-I error?

About 50 of them, give or take some for chance error.

None of them.

All of them.

About 100 of them, give or take some for chance error.

About half of them, give or take some for chance error.

Types of Error: E

In a certain population, the mean mu is 45. There are one thousand statisticians, and none of them know what mu is. Each statistician takes a simple random sample from this population, and each one uses his/her sample to perform the following two-sided test:

H0: mu = 42

Ha: mu != 42

How many of them will commit a Type-I error?

About 50 of them, give or take some for chance error.

None of them.

All of them.

About 100 of them, give or take some for chance error.

About half of them, give or take some for chance error.

Aspirin I: R-Code

In the `abd` package we find the data frame `Aspirin`:

```
data(Aspirin)
View(Aspirin)
help(Aspirin)
```

This data frame presents the results of an experiment with a completely randomized design. Researchers wanted to know whether taking aspirin in low doses affects one's risk of developing cancer. They had 39,876 subjects willing to participate in the experiment, and they randomly divided the subjects into two groups, indicated by the variable **treatment** with values "Aspirin" and "Placebo". Subjects in the Aspirin group took aspirin each day for a long period of time. Subjects in the Placebo group took a substance that looks and tastes like aspirin, but which has no effect on the body. After the study ended, researchers recorded whether or not each subject had developed cancer, as indicated in the variable **cancer** (yes, no).

Researchers decided to investigate their question by performing a two-sided test of significance. The parameters of interested were defined as follows:

Define:

p1 = the proportion of ALL individuals who would develop cancer, if all of them were to take aspirin like the subjects in the Aspirin group did.

p2 = the proportion of ALL individuals who would develop cancer, if all of them were to take a placebo, like the subjects in the placebo group did.

The hypotheses were:

H0: p1 = p2 (aspirin has no affect on risk for cancer)

Ha: p1 != p2 (aspirin affects the risk of cancer)

The researchers also decided that they would like a 95%-confidence interval for p1 - p2.

Help the researchers: which of the following bit of R-code will give the them what they want?

```
ttestGC(cancer~treatment,data=Aspirin,mu=0)
```

Feedback: ttestGC() is used when the response variable is numerical and when we are interested in means. Here the response variable **cancer** is a factor variable with two values (yes, no). We are counting up yesses, and are interested in the proportion of yesses. R will throw an error if you give it this code.

```
proptestGC(cancer~treatment,data=Aspirin,
           p=0,success="yes",
           first="Aspirin")
```

Feedback: Remember that both of your variables **treatment** and **cancer** are factor variables. The formula you've got here looks like one that you would use when you have a *numerical* response variable. R will throw an error an error if given this code.

```
proptestGC(~treatment+cancer,data=Aspirin,
           p=0,success="yes",
           first="Placebo")
```

Feedback: Close, but the way the researchers defined their parameters, the Aspirin group needs to come first.

```
proptestGC(~cancer+treatment,data=Aspirin,
           p=0,success="yes",
           first="Aspirin")
```

Feedback: Close, but you switched the variables around. You need to put your explanatory variable (**treatment**) first.

```
proptestGC(~treatment+cancer,data=Aspirin,
           p=0,success="yes",
           first="Aspirin")
```

Feedback: Just what the doctor ordered!

Aspirin II: Safety Considerations

Regarding the experiment in the data frame `Aspirin` from the `abd` package, the researchers wanted to know whether or not taking aspirin affects one's risk of developing cancer. Recall that they defined their parameters as follows:

p1 = the proportion of ALL individuals who would develop cancer, if all of them were to take aspirin like the subjects in the Aspirin group did.

p2 = the proportion of ALL individuals who would develop cancer, if all of them were to take a placebo, like the subjects in the placebo group did.

They ran the code for a two-sided significance test and got the following results:

```
##
##
## Inferential Procedures for the Difference of Two Proportions p1-p2:
##    cancer grouped by treatment
##
##
## Descriptive Results:
##
##          yes     n estimated.prop
## Aspirin 1438 19934        0.07214
## Placebo 1427 19942        0.07156
##
##
## Inferential Results:
##
## Estimate of p1-p2:    0.0005805
## SE(p1.hat - p2.hat):  0.002586
##
## 95% Confidence Interval for p1-p2:
##
##          lower.bound          upper.bound
##          -0.004489            0.005650
##
## Test of Significance:
##
```

```
##  H_0:  p1-p2 = 0
##  H_a:  p1-p2 != 0
##
##  Test Statistic:     z = 0.2245
##  P-value:        P = 0.8224
```

Regarding the "safety" of applying `porptestGC()` to the situation, which of the following remarks are good to make? (There are two correct answers.)

Option A:

The researchers defined their parameters so as to refer to the population of ALL people. If they really want the results of their experiment to apply to this population, they should good grounds for believing that the subjects in the experiment are a representative sample from the population.

Option B:

The subjects were a simple random sample from the population, so it's definitely safe to use `proptestGC()`.

Option C:

In order to make confidence intervals and P-values, `proptestGC()` uses the normal-curve approximation. This approximation may not be good unless there are at least ten successes and at least ten failures in each sample. This time we have thousands of successes and thousands of failures in each sample. We also did a randomized experiment, so in all relevant respects it is safe to use `proptestGC()`.

Option D:

It's not safe to use `proptestGC()` because the subjects were not a simple random sample from the population.

Aspirin III: Decision and Conclusion

Regarding the experiment in the data frame `Aspirin` from the `abd` package, the researchers wanted to know whether or not taking aspirin affects one's risk of developing cancer. Recall that they defined their parameters as follows:

p1 = the proportion of ALL individuals who would develop cancer, if all of them were to take aspirin like the subjects in the Aspirin group did.

p2 = the proportion of ALL individuals who would develop cancer, if all of them were to take a placebo, like the subjects in the placebo group did.

They ran the code for a two-sided significance test and got the following results:

```
##
##
## Inferential Procedures for the Difference of Two Proportions p1-p2:
##   cancer grouped by treatment
##
##
## Descriptive Results:
##
##          yes     n estimated.prop
## Aspirin 1438 19934        0.07214
## Placebo 1427 19942        0.07156
##
##
## Inferential Results:
```

```
## 
## Estimate of p1-p2:    0.0005805
## SE(p1.hat - p2.hat):  0.002586
## 
## 95% Confidence Interval for p1-p2:
## 
##          lower.bound         upper.bound
##          -0.004489           0.005650
## 
## Test of Significance:
## 
##  H_0:  p1-p2 = 0
##  H_a:  p1-p2 != 0
## 
##  Test Statistic:     z = 0.2245
##  P-value:        P = 0.8224
```

Which of the following options are correct? (Two are correct: one for the decision, and one for the conclusion. Find and select both of them.)

Option A:

Since the P-value is 0.8224, which is bigger than our usual cut-off of 0.05, we fail to reject the Null Hypothesis.

Option B:

This data provided very little evidence that using aspirin has any effect on one's risk of developing cancer.

Option C:

Since the P-value is 0.8224, which is bigger than our usual cut-off of 0.05, we reject the Null Hypothesis.

Option D:

This data provided strong evidence that aspirin affects one's risk of developing cancer.

Option E:

This data provides strong evidence that aspirin has NO effect on one's risk of developing cancer.

Feedback: The way we set up tests, a high P-value does not provide evidence FOR the Null. It only means that the data did not provide much evidence AGAINST it.

Option F:

7.214% of the aspirin users developed cancer, whereas only 7.156% of the members of the placebo group developed cancer. Hence we can conclude that using aspirin slightly increases one's risk of developing lung cancer.

Feedback: The stated percentages are correct, but the whole point of the test of significance is to see whether the difference between those two percentages could be reasonably explained as due just to chance variation in the random assignment of subjects to treatment groups. As the high P-value shows, the observed difference could easily be due just to chance.

One Proportion, Summary Data: I

It is known that 60% of all Ohio State students plan to vote for the Democratic candidate in the next presidential election. In a survey of 200 randomly-selected Georgetown College students, it is found that 101 of them plan to vote for the Democratic candidate.

The Research Question is:

Is the level of support for the Democratic candidate lower at GC than it is at OSU?

Let's define:

p = the proportion of ALL GC students who would vote for the Democratic candidate.

We've got some prior evidence that GC students are a bit on the conservative side, compared to students at most other colleges, so we decide to perform a one-sided test of significance:

H0: p = 0.60 (GC students are like OSU students, as far as voting is concerned)

Ha: p < 0.60 (GC students are less favorable to the Democratic candidate than OSU students are)

What's the right R-code to run for this test?

```
binomtestGC(x=101,n=200,p=0.60,
            alternative="less")
```

Feedback: Correct!

```
binomtestGC(x=101,n=200,p=0.60)
```

Feedback: Oops, this gives a TWO-sided test.

```
binomtestGC(x=101,n=200,p=0,
            alternative="less")
```

Feedback: This says that the Null Hypothesis says p = 0. But the Null says p = 0.60, right?

```
binomtestGC(x=101,n=200)
```

Feedback: This gives you only a confidence interval, not the test.

One Proportion, Summary Data: II

It is known that 60% of all Ohio State students plan to vote for the Democratic candidate in the next presidential election. In a survey of 200 randomly-selected Georgetown College students, it is found that 101 of them plan to vote for the Democratic candidate.

The Research Question is:

Is the level of support for the Democratic candidate lower at GC than it is at OSU?

Let's define:

p = the proportion of ALL GC students who would vote for the Democratic candidate.

We've got some prior evidence that GC students are a bit on the conservative side, compared to students at most other colleges, so we decide to perform a one-sided test of significance:

H0: p = 0.60 (GC students are like OSU students, as far as voting is concerned)

Ha: p < 0.60 (GC students are less favorable to the Democratic candidate than OSU students are)

Researchers perform the test, and they get the following results:

```
## 
## Exact Binomial Procedures for a Single Proportion p:
##  Results based on Summary Data
## 
## 
## Descriptive Results:  101 successes in 200 trials
## 
## Inferential Results:
## 
## Estimate of p:    0.505
## SE(p.hat):    0.0354
## 
## 95% Confidence Interval for p:
## 
##          lower.bound        upper.bound
##          0.000000           0.565300
## 
## Test of Significance:
## 
##  H_0:  p = 0.6
##  H_a:  p < 0.6
## 
##  P-value:        P = 0.004
```

What's the best way to interpret the P-value in this test?

Option A:

There is about a 0.4% chance that the p = 0.60 and the Null is correct.

Feedback: The Null is either right or not, no "chances" about it. P-values don't give the chance that the Null is right.

Option B:

If 60% of all GC students would vote Democratic, then there is about a 0.4% chance of getting 101 or fewer students (out of 200 sampled) saying that they would vote Democratic.

Feedback: Good job. Note especially the one-sided language ("101 or fewer . . .) used in the interpretation.

Option C:

If less than 60% of all GC students would vote Democratic, then there is about a 0.4% chance of getting 101 or fewer students (out of 200 sampled) saying that they would vote Democratic.

Feedback: Careful. P-value interpretations start off saying something like: "If the Null is RIGHT, then . . . ".

Option D:

If the Null is right, then there is a 0.004% chance of getting results at least as far below what the Null expects as the results we actually got were.

Feedback: This is almost right, except for an arithmetic mistake: 0.004 corresponds to 0.%$, not to 0.004%.

One Proportion, Summary Data: III

It is known that 60% of all Ohio State students plan to vote for the Democratic candidate in the next presidential election. In a survey of 200 randomly-selected Georgetown College students, it is found that 101 of them plan to vote for the Democratic candidate.

The Research Question is:

Is the level of support for the Democratic candidate lower at GC than it is at OSU?

Let's define:

p = the proportion of ALL GC students who would vote for the Democratic candidate.

We've got some prior evidence that GC students are a bit on the conservative side, compared to students at most other colleges, so we decide to perform a one-sided test of significance:

H0: p = 0.60 (GC students are like OSU students, as far as voting is concerned)

Ha: p < 0.60 (GC students are less favorable to the Democratic candidate than OSU students are)

Researchers perform the test, and they get the following results:

```
##
## Exact Binomial Procedures for a Single Proportion p:
##  Results based on Summary Data
##
##
## Descriptive Results:   101 successes in 200 trials
##
## Inferential Results:
##
## Estimate of p:     0.505
## SE(p.hat):     0.0354
##
## 95% Confidence Interval for p:
##
##          lower.bound          upper.bound
##          0.000000             0.565300
##
## Test of Significance:
##
##  H_0:  p = 0.6
##  H_a:  p < 0.6
##
##  P-value:         P = 0.004
```

Which of the following sets of remarks is best?

Option A:

The P-value is less than 0.05, so we reject H0. This study provided strong evidence that GC students are less likely to favor the Democratic candidate than OSU students are.

Feedback: Good job on decision and conclusion!

Option B:

The p-value is less than 0.05, so we do not reject the Null Hypothesis. This study provided strong evidence that GC students are less likely to favor the Democratic candidate than OSU students are.

Option C:

The P-value is less than 0.05, so we reject H0. This study did not provide strong evidence that GC students are less likely to favor the Democratic candidate than OSU students are.

Option D:

The P-value is less than 0.05, so we do not reject H0. This study did not provide strong evidence that GC students are less likely to favor the Democratic candidate than OSU students are.

Comparing Two Groups

A random sample of 200 students from the University of Wisconsin has 120 of them planning to vote for the Democratic candidate in the next Presidential election. An independent random sample of 200 Georgetown College students has 101 of them planning to vote for the Democratic Candidate.

We would like to know which campus is more favorable to Democrats. We have no prior evidence to suggest an answer one way or another, so we decide to perform a two-sided test of significance.

Figure out on your own how to define parameter(s) and state hypotheses, and decide what R-code you should run. Then study the output and choose the best answer below.

Option A:

The P-value is bigger than 0.05, so we do not reject H0. This study did not provide sufficiently strong evidence to conclude that the two campuses differ in their favor ability towards Democrats.

Feedback:

Right. Hopefully you used either this code:

```
proptestGC(x=c(120,101),n=c(200,200),p=0)
```

or this code:

```
proptestGC(x=c(101,120),n=c(200,200),p=0)
```

depending on which group you wanted to count as "first".

Option B:

The P-value is bigger than 0.05, so we reject H0. This study did not provide sufficiently strong evidence to conclude that the two campuses differ in their favorability towards Democrats.

Option C:

The P-value is less than 0.05, so we do not reject H0. This study provided sufficiently strong evidence to conclude that the two campuses differ in their favorability towards Democrats.

Option D:

The P-value is less than 0.05, so we reject H0. This study provided sufficiently strong evidence to conclude that the two campuses differ in their favorability towards Democrats.