Fall 2019 MSCR 500 & 533 Quiz 2

Anish Shah

11/27/19

Question 1

Description

Study of long-term effects of eye surgery to correct myopia
44 pts receive surgery in pilot study
310 pts as part of core study get surgery
Total of 354 pts
Follow up of 10 years, with measure of refractive error collected

Is the long-term change in refractive error different for the pilot and core study group? Use alpha = 0.10

Data from SAS output

Pilot patients:

n = 44
Mean = 0.7443
SD = 1.02829
Var = 1.0574

Core patients:

N = 310
Mean = 0.7174
SD = 1.0613
Var = 1.1263

Answer to Question 1

Hypothesis

H0: u1 - u2 = 0
H1: u1 - u2 =/= 0
u1 = mean refractive change of pilot group
u2 = mean refractive change of core group

Test: two sample t-test (unpooled)

This is the most appropriate test because we have two normally distributed (by CTL) independent populations, where we are comparing the mean of two different populations. Also, the distributions appear to be both symmetrical and bellshaped. However, the literature suggests that unless the populations are the same size, its more appropriate to assume that the variance is unequal. Thus, we will use the unpooled t-tes t.

Test Statistic

T = (xbar1 - xbar2) - 0 / sqrt(s1^2/n1 + s2^2/n2)
= 0.0269 / 0.1663287
= 0.1617

P-value

T ~ T(43) distribution (by Satterwaite)
P(T = 0.1617) < 1.302 (calculated by T-table, where our T-statistic is much lower than the threshold T-stat for an alpha = 0.10)

Fail to Reject H0 at alpha = 0.10

Based on the given parameters, we have insufficient evidence to reject the null hypothesis, and can conclude that with an alpha = 0.10, there is no difference between the mean change in refraction error over 10 years in the pilot and core groups.

Question 2

Description

Students get a class pre-test and post-test

Question: Was there an improvement in test scores?

Data from SAS output

N = 20
Mean = -12
SD = 21.9089
Var = 480

Answer to Question 2

Hypothesis:

H0 = u = 0
H1 = u > 0
u = true average difference between pre-test and post-test scores

Test: paired t-test

This test is most appropriate (the one-sample paired t-test) because it is a population that is tested twice (e.g. before/after intervention), thus each result is "paired" between groups.

Test statistic

T = xbar - 0 / (sigma / sqrt(n))
= -12 / 4.898979
= -2.4495

P-value

T ~ t(19) under H0
P(T = -2.4495) <= 2.093 of t(0.025)

Reject H0 at alpha = 0.05

As the T-statistic falls between t(0.025) and t(0.01), we can say that our p-value is ~ 0.025. Thus, we have enough evidence to reject the null hypothesis and conclude that the scores were lower in the post-test than in the pre-test. Hopefully this is not our class.

Question 3

Answer to question 3

Hypothesis:

H0: u1 - u2 = 0
H1: u1 - u2 =/= 0
u1 = mean calories of frozen food
u2 = mean calories of fresh food

Test: two sample t-test (unpooled)

This is the most appropriate test because we have two normally distributed (by CTL) independent populations, where we are comparing the mean of two different populations. Also, the distributions appear to be both symmetrical and bellshaped. However, the literature suggests that unless the populations

are the same size, its more appropriate to assume that the variance is unequal. Thus, we will use the unpooled t-tes t.

Test statistic

T = 7.81

P-value

P < 0.0001

Reject H0 at alpha = 0.01

Based on the t-statistic and P-value, we have enough evidence to reject the null hypothesis and accept that there is a difference in the calories in frozen versus fresh food.

Probability Distribution

T ~ T(df = 61.627) <- as calculated by the Satterthwaite formula by SAS
I used this because the variances appear to be different between populations

99% Confidence Intervals

The 99% confidence interval around the difference in mean between frozen and fresh calories is mean = 194.8 (66.3 - 323.3).

My SAS code

```
/*
SAS code for Quiz 2
Anish Shah
11/27/19
*/

*Set library;
LIBNAME H "H:\My Documents\Github\tl1\mscr\biostats\";

* Import data;
PROC IMPORT OUT = WORK.sandwich_name
    DATAFILE = "H:\My Documents\Github\tl1\mscr\biostats\sandwich_name.xls";
    GETNAMES = YES;
    MIXED = YES;
    SCANTEXT = YES;
    USEDATE = YES;
    SCANTIME = YES;
RUN;


PROC IMPORT OUT = WORK.sandwich_nutrition
```

```
    DATAFILE = "H:\My Documents\Github\tl1\mscr\biostats\sandwich_nutrition.xls";
    GETNAMES = YES;
    MIXED = YES;
    SCANTEXT = YES;
    USEDATE = YES;
    SCANTIME = YES;
RUN;

*Examine contents of data sets;
PROC CONTENTS DATA = WORK.sandwich_name; RUN;
PROC CONTENTS DATA = WORK.sandwich_nutrition;
RUN;

* Merge into temp file by sandwich_number into temp data set called sandwiches;
PROC SORT DATA = WORK.sandwich_name;
    BY Sandwich_Number;
RUN;

PROC SORT DATA = WORK.sandwich_nutrition;
    BY Sandwich_Number;
RUN;

DATA WORK.sandwiches;
    MERGE WORK.sandwich_name WORK.sandwich_nutrition;
    BY Sandwich_Number;
RUN;

* Descriptive analysis;
PROC UNIVARIATE DATA = WORK.sandwiches;
    VAR Calories TFat Protein Carb Fiber Sodium Weight;
    HISTOGRAM Calories TFat Protein Carb Fiber Sodium Weight;
    PROBPLOT Calories TFat Protein Carb Fiber Sodium Weight;
RUN;

PROC FREQ DATA = WORK.sandwiches;
    TABLES Brand Name Category Frozen;
RUN;

* Modify teh data... change category of BEEF to Beef;
DATA WORK.sandwiches;
    SET WORK.sandwiches;
    * Decapitate the cow;
    IF Category = "BEEF" THEN Category = "Beef";
RUN;

PROC CONTENTS DATA = WORK.sandwiches;
```

```
RUN;

* Test average calories different for fresh or frozen, alpha = 0.01;
PROC TTEST DATA = WORK.sandwiches ALPHA=0.01;
    CLASS Frozen;
    VAR Calories;
    TITLE "Difference in calories for fresh or frozen sammies";
RUN;

* 99% confinddence in difference in mean calories of frozen versus fresh sammies;
PROC MEANS DATA = WORK.sandwiches CLM ALPHA = 0.01;
    CLASS Frozen;
    VAR Calories;
RUN;
```