PS6

2024-10-15

Stat 5000 Homework #4
Fall 2024 due Fri, October 18th @ 11:59 pm Name:
Collaborators: Ethan, **The Hatman**

Q1

Consider the nesting example from the previous homework assignment. The areas of entrances to nesting cavities were measured for 294 nesting sites for nine common species of birds and rodents in Oregon. Sample sizes, sample means and sample standard deviations are shown below with samples means and sample standard deviations computed from the natural logarithms of the observed areas of cavity entrances.

Summary Statistics for Natural Logarithm of Areas of Nesting Cavity Entrances

Species	*Sample Size (n)*	*Sample Mean Log(1000 mm²)*	*Sample Std. Dev. Log(1000 mm2)*
Mouse	127	7.347	0.4979
Pinyon mouse	44	7.369	0.4235
Bewick's wren	24	7.428	0.3955
Mountain bluebird	41	7.487	0.3181
Ash-throated flycatcher	18	7.563	0.3111
Plain titmouse	16	7.568	0.4649
Northern flicker	11	8.214	0.2963
Western Screech-owl	7	8.272	0.3242
American kestrel	6	8.297	0.5842

Figure 1: img 1

Previously, we computed the ANOVA table and found that at least one of the species had a different mean log-area of nesting cavity entrances.

Variation	DF	SS	MS	F
Model	8	17.420	2.178	11.350
Error	85	54.696	0.1919	p_{val}
Total	293	72.116		≈ 0

After observing the data, the researchers decided to test the null hypothesis that the average of the logarithm of the area of cavity openings for the first six species (mouse, pinyon mouse, Bewick's wren, mountain bluebird, ash-throated flycatcher and plain titmouse) is equal to the average of the logarithm of the cavity opening for the other three species (northern flicker, western screech owl, and American kestrel).

$$H_O : \gamma = \frac{\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6}{6} - \frac{\mu_7 + \mu_8 + \mu_9}{3} = 0$$

There is no file with the actual data. You will need to answer the following questions using only the summary statistics shown in the tables. You may also use the following quantiles:

$$t_{N-r,1-\alpha/2} = 1.973$$
$$t_{N-r,1-\alpha/(2m)} = 3.246$$
$$F_{r-1,N-r,1-\alpha} = 1.989$$
$$F_{1,N-r,1-\alpha/2} = 5.107$$

(a)

Because the researchers decided to perform this test after examining the summary statistics and selecting what appeared to be a large difference, they will need to use the Scheffe method. Report (i) the test statistic and (ii) the critical value. (iii) Interpret the result in the context of the study.

```
y_i <- (7.347 + 7.369 + 7.428 + 7.487 + 7.563 + 7.568) / 6
y_j <- (8.214 + 8.272 + 8.297) / 3
testStat <- abs(y_i - y_j)
n_i <- 127 + 44 + 24 + 41 + 18 + 16
n_j <- 11 + 7 + 6
frac <- (1/n_i) + (1/n_j)
seconTerm <- sqrt(0.1919 * frac)
firstTerm <- sqrt(8 * 1.989)
compStat <- firstTerm * seconTerm
testStat >= compStat
```

```
## [1] TRUE
```

$$|\bar{Y}_i - \bar{Y}_j| \geq \sqrt{(r-1)F_{r-1,N-r,1-\alpha}}\sqrt{MS_{error}(\frac{1}{n_i} + \frac{1}{n_j})}$$

(i): Test Statistic: 0.80067

(ii): Critical value: 0.37221

(iii): Interpretation

Test Statistics $\geq$ Critical Value $\rightarrow$ then significant!

$$0.8006667 \geq 0.3722083$$

(b)

Use the Scheffe method to construct a confidence interval for the contrast γ such that the confidence level is at least 95 percent.

Typical procedure:

$$\hat{\gamma} \pm t_{N-r,1-\alpha/2} S_{\hat{\gamma}}$$

Where:

$$S_{\hat{\gamma}} = \sqrt{MS_{error} \sum_i (c_i^2/n_i)}$$

$$t_{N-r,1-\alpha/2} = 1.973$$

However, for Scheffe, we use

$$\sqrt{(r-1)F_{r-1,N-r,1-\alpha}}$$

Instead of

$$t_{N-r,1-\alpha/2} = 1.973$$

Where:

$$F_{r-1,N-r,1-\alpha} = 1.989$$

```
# sum for all i c_i^2 / n_i
plus <- ((1/6)^2)/127 + ((1/6)^2)/44 + ((1/6)^2)/24 + ((1/6)^2)/41 + ((1/6)^2)/18 + ((1/6)^2)/16

minus <- ((-1/3)^2)/11 + ((-1/3)^2)/7 + ((-1/3)^2)/6

coef <- plus + minus
coef
```

```
## [1] 0.05045682
```

```
plusminus <- sqrt((9-1) * 1.989) * sqrt(0.1919 * coef)
plusminus
```

```
## [1] 0.3925182
```

```
y_i <- (7.347 + 7.369 + 7.428 + 7.487 + 7.563 + 7.568) / 6
y_j <- (8.214 + 8.272 + 8.297) / 3
estimate <- y_i - y_j

estimate + plusminus
```

```
## [1] -0.4081485
```

```
estimate - plusminus
```

```
## [1] -1.193185
```

We are $100(1-\alpha)$ confident that the true value of the contrast γ is between -0.4081 and -1.1932.

(c)

Construct a confidence interval for the contrast γ using the formula based on an ordinary t-test without any multiple comparison adjustment. How does it compare to the result from the Scheffe method?

Estimate of the contrast, γ, remains the same, we just use a different adjustment in the plus/minus:

```
# sum for all i c_i^2 / n_i
plus <- ((1/6)^2)/127 + ((1/6)^2)/44 + ((1/6)^2)/24 + ((1/6)^2)/41 + ((1/6)^2)/18 + ((1/6)^2)/16

minus <- ((-1/3)^2)/11 + ((-1/3)^2)/7 + ((-1/3)^2)/6

coef <- plus + minus
coef
```

```
## [1] 0.05045682
```

```
plusminus <- 1.973 * sqrt(0.1919 * coef)
plusminus
```

```
## [1] 0.1941442
```

```
y_i <- (7.347 + 7.369 + 7.428 + 7.487 + 7.563 + 7.568) / 6
y_j <- (8.214 + 8.272 + 8.297) / 3
estimate <- y_i - y_j

estimate + plusminus
```

```
## [1] -0.6065224
```

```
estimate - plusminus
```

```
## [1] -0.9948109
```

We are $100(1-\alpha)$ confident that the true value of the contrast γ is between -0.6065224 to -0.9948109 using the ordinary t-test method.

This confidence interval is slightly narrower (smaller) than the one calculated using Scheffe's Method.

(d)

One researcher tried to argue that using the confidence interval from (c) in situations such as this is okay because only one confidence interval was constructed. How would you respond to this assertion?

This is not ok. This is actually fairly problematic. We have already done a number of post-hoc transformations and calculations to get at where we started in this problem, from transforming the data to the log scale to normalize our data to even the construction/parametrization of the contrast in question. In a sense, we have already done "multiple comparisons", so it would be disingenuous for us to use the single confidence interval they are interested in. It would be more appropriate to use an adjusted one that accounts for multiple comparisons, a la Scheffe's Method; despite this method being more conservative in its estimate (wider interval than the ordinary t-test), it's worth noting that this difference isn't even especially large!

Q2

In 1879, A. A. Michelson made 100 determinations of the velocity of light in air. The data used in this analysis were reported by Stigler (1977, Annals of Statistics, 5:4, 1075). The currently accepted "true" velocity of light in a vacuum is 299,792.5 km/sec, but the velocity of light in air could be slower. The data for this exercise were modified by Stigler to correct for overall bias in Michelson's measurement for the speed of light in air (i.e. the numbers are in km/sec, and have had 299,000 subtracted from them).

The measurements are grouped into five trials with 20 determinations for each trial. Since each determination was an attempt to measure the same "true" value of the speed of light in air, one might expect that the population means of possible measurements should be the same for all five trials. However, adjustments to the equipment or method for measuring the speed of light in air may have been made between trials, and these may cause the mean values to differ across trials. (Note that although one can argue that the trails are random effects, for this problem, let's treat trials as fixed effects.) The data is located in the file lightspeed.csv saved in our course's shared folder in SAS Studio.

(a)

Use the Tukey-HSD method in SAS to compare means for measurements of speed of light in air for each pair of trials

The GLM Procedure

Tukey's Studentized Range (HSD) Test for BiasAdj

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	95
Error Mean Square	5510.632
Critical Value of Studentized Range	3.93274
Minimum Significant Difference	65.28

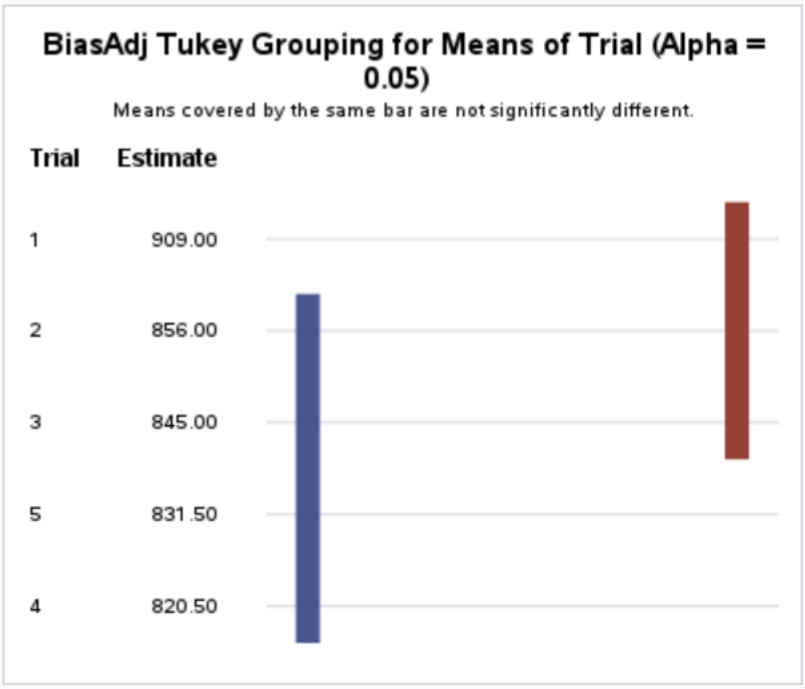


Figure 2: img

i.

What is the minimum distance between two sample means required to reject the hypothesis that two trials have the same means using an experiment-wise type I error level of $\alpha = 0.05$?

65.28

ii.

Indicate which means are significantly different at the 5% level.

Trial 1 is significantly different than Trial 4 and Trial 5 at the 5% level. Likewise, Trial 4 is different from Trial 1 and Trial 4 is different than Trial 5 at the 5% level.

(b)

If we are interested in using Bonferroni's method to perform a hypothesis test for differences between all pairs of means, what is the type I error rate that we need to control for each comparison so that the experiment-wise type I error rate is 0.05?

The GLM Procedure

Bonferroni (Dunn) t Tests for BiasAdj

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	95
Error Mean Square	5510.632
Critical Value of t	2.87407
Minimum Significant Difference	67.468

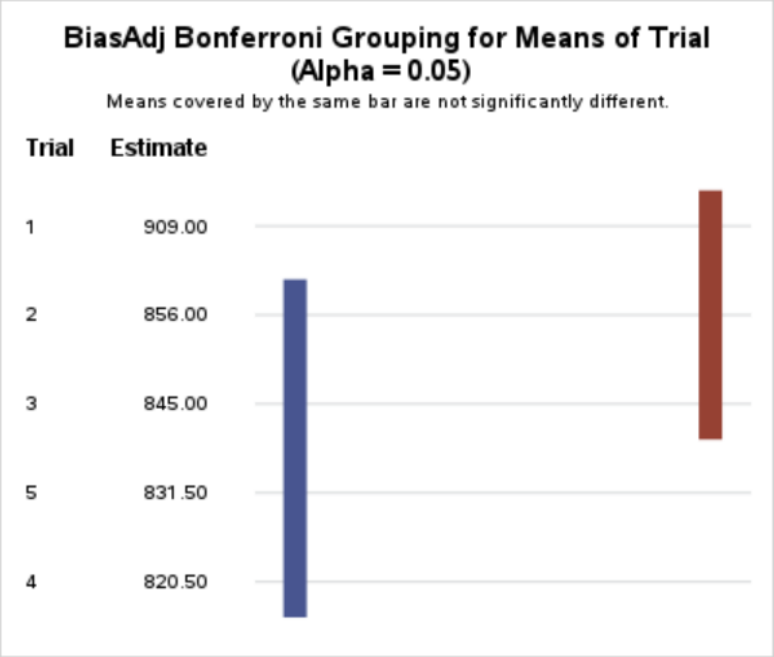


Figure 3: img

For m tests, we use α/m instead of α. For $\alpha = 0.05$, $m = 10 \equiv \binom{5}{2}$, we have

Type 1 Error Rate: 0.005.

(c)

Consider the following set of orthogonal contrasts that examine potential chronological changes in the population means across the five trials:

Contrast 1: (mean for the first trial) - (mean of the other four trials) *Contrast 2:* (mean for the second trial) - (mean of the last three trials) *Contrast 3:* (mean for the third trial) - (mean of the last two trials) *Contrast 4:* (mean for the fourth trial) - (mean of the last trials)

i.

For each contrast, report the contrast coefficients, the sum of squares, corresponding F-statistic, and p-value

Contrast 1: Contrast coefficients: 70.75 Sum of squares: 80089 F-statistic: 14.53 p-value: 0.0002

Contrast 2: Contrast coefficients: 23.667 Sum of squares: 8401.667 F-statistic: 1.52 p-value: 0.220

Contrast 3: Contrast coefficients: 19 Sum of squares: 4813.333 F-statistic: 0.87 p-value: 0.3524

Contrast 4: Contrast coefficients: -11 Sum of squares: 1210 F-statistic: 0.22 p-value: 0.6404

The GLM Procedure

Dependent Variable: BiasAdj

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	94514.0000	23628.5000	4.29	0.0031
Error	95	523510.0000	5510.6316		
Corrected Total	99	618024.0000			

R-Square	Coeff Var	Root MSE	BiasAdj Mean
0.152929	8.708779	74.23363	852.4000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Trial	4	94514.00000	23628.50000	4.29	0.0031

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Trial	4	94514.00000	23628.50000	4.29	0.0031

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Contrast1	1	80089.00000	80089.00000	14.53	0.0002
Contrast2	1	8401.66667	8401.66667	1.52	0.2200
Contrast3	1	4813.33333	4813.33333	0.87	0.3524
Contrast4	1	1210.00000	1210.00000	0.22	0.6404

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Contrast1	70.7500000	18.5584071	3.81	0.0002
Contrast2	23.6666667	19.1670404	1.23	0.2200
Contrast3	19.0000000	20.3297164	0.93	0.3524
Contrast4	-11.0000000	23.4747345	-0.47	0.6404

Figure 4: img

ii.

Using a Type I error level of 0.05 for each test in part (i), summarize the test results. (State all interpretations in the context of the study.)

We have evidence to reject the null hypothesis at the $\alpha = 0.001$ level that the first contrast is zero. This is to say we have evidence to support the alternative hypothesis that the first contrast is not zero, which is to say we have evidence that the mean for the first trial is different from the means of the other four trials.

However, the statistical tests associated with the other 3 contrasts are the opposite; we do not have significant evidence to reject the null hypothesis of contrasts 2, 3, and 4, which is to say we do not have evidence to reject there being a difference in the mean of the second trial compared with the average of the other 3 trials, nor of there being a difference between the average of the third trial and the average of the last 2 trials, nor of there being a difference between the mean of the fourth trial and the average of the last trial.

iii.

Show that the sums of squares for the set of orthogonal contrasts sums to the between trials sums of squares (SS_{model}).

From our overall model summary:

$$SS_{model} = 94514$$

Taken with the information above, we have:

$$SS_{\text{contrast 1}} + SS_{\text{contrast 2}} + SS_{\text{contrast 3}} + SS_{\text{contrast 4}} = 80089 + 8401.667 + 4813.333 + 1210 = 94514$$

Such that

$$94514 = SS_{model} = SS_{\text{contrast 1}} + SS_{\text{contrast 2}} + SS_{\text{contrast 3}} + SS_{\text{contrast 4}}$$

Q3

Ten patients who suffered from a sleep disorder were examined in a study of the effectiveness of a sleep-aid drug. First, each patient was given a placebo, a pill that did not contain any active ingredient. The average number of hours of sleep for three nights was recorded for each patient. Then, each patient was given a pill containing the drug for three consecutive nights and average hours of sleep was recorded for each patient. The data appear below (and can also be found in the sleep.csv file saved in our course's shared folder on SAS Studio).

	Hours of Sleep per Night	
Patient	Drug Given	Placebo Given
1	1.3	0.6
2	1.1	1.1
3	6.2	2.5
4	3.6	2.8
5	4.9	2.9
6	1.4	3.0
7	6.6	3.2
8	4.5	4.7
9	4.3	5.5
10	6.1	6.2

Figure 5: img 2

(a)

Identify the blocks and treatments for this experiment.

Block is an individual suffering from a sleep disorder. Treatments are receiving either a placebo or a sleep-aid drug.

(b)

Explain why this is a matched pairs experiment.

Each block (an individual) effectively provides multiple measurements in our data, specifically for the two different variables of interest (Placebo and Drug). This is a reuse matched pairs experiment.

(c)

What is one aspect of the experimental design that could be improved? Explain how this would strengthen the results.

Additionally, more data would be helpful (low hanging fruit); so either having more participants, or having more nights under study could help. A third one I'll throw in for good measure: Having different dosages of the sleeping pill could also strengthen results, as it could affirm the impact of the sleep-aid drug while also possibly quantifying dosage effects.

(d)

Using SAS, compute the value of a t-statistic for testing the null hypothesis that mean hours of sleep are the same for the drug and the placebo, against a one-sided alternative that mean hours of sleep are greater when the drug is used. Report the (i) test statistic, (ii) degrees of freedom, (iii) p-value, and (iv) interpret the result in the context of the study.

tTest Drug - Placebo Sides Upper

The TTEST Procedure

Variable: diff

N	Mean	Std Dev	Std Err	Minimum	Maximum
10	0.7500	1.7890	0.5657	-1.6000	3.7000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0.7500	-0.2871	Infty	1.7890	1.2305	3.2660

DF	t Value	Pr > t
9	1.33	0.1088

Figure 6: img

Test Statistic: 1.33 degrees of freedom: 9 p-value: 0.1088 Interpretation:

We have evidence at the $\alpha = 0.20$ level to reject the null hypothesis that the average sleep times of users of the sleep-aid drug is less than or equal to the average sleep times of users of the placebo. Typically, for most $\alpha = 0.10, 0.05, 0.01, etc.$ levels we do not reject the null hypothesis stated previously. Overall, we would say that we do not reject the null hypothesis and support there being evidence that the average sleep times of users of the sleep-aid drug is less than or equal to the average sleep times of users of the placebo.

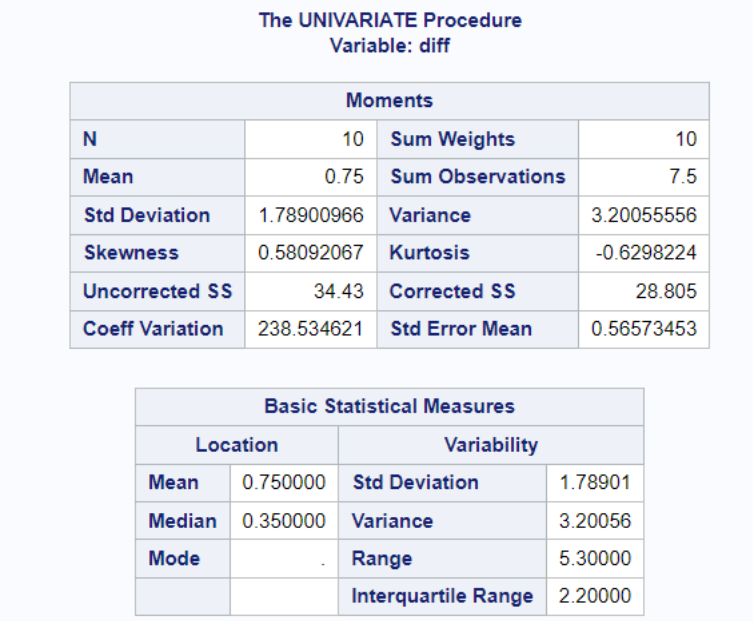
(e)

Check the assumptions for this test and provide supporting data from SAS.

Our assumptions are independence within groups, i.e. independence between blocks (not between because of matched pairs!), and normality of differences. Overall, these assumptions do not readily appear to be violated.

As far as independence between blocks, I do not believe we have reason to suspect this is being violated. It would be violated if participants were sleeping with one another, or were sleeping in an environment where they could disturb another participant. But this does seem a bit farcical.

The below provides a number of diagnostics run for normality. Overall, the statistical tests provide evidence not to reject the null hypothesis of normality, particularly Shapiro-Wilk failing to reject normality. Furthermore, the box plot and histograms tend to have a normal shape, though the box plot does show that variance (spread) is not symmetrical. Furthermore, skewness and (excess) kurtosis both have absolute differences less than 1, and while mean and median are not equal, they are relatively similar (difference of 0.4 for values having a range of 5.3). Additionally, the quantile (qq) plot shows values that closely track with the reference line, though there is some inexactness.

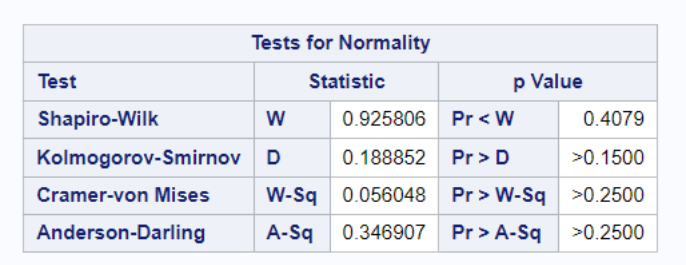


The UNIVARIATE Procedure
Variable: diff

Moments			
N	10	Sum Weights	10
Mean	0.75	Sum Observations	7.5
Std Deviation	1.78900966	Variance	3.20055556
Skewness	0.58092067	Kurtosis	-0.6298224
Uncorrected SS	34.43	Corrected SS	28.805
Coeff Variation	238.534621	Std Error Mean	0.56573453

Basic Statistical Measures			
Location		Variability	
Mean	0.750000	Std Deviation	1.78901
Median	0.350000	Variance	3.20056
Mode	.	Range	5.30000
		Interquartile Range	2.20000

Figure 7: img 2



Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.925806	Pr < W	0.4079
Kolmogorov-Smirnov	D	0.188852	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.056048	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.346907	Pr > A-Sq	>0.2500

Figure 8: img 2

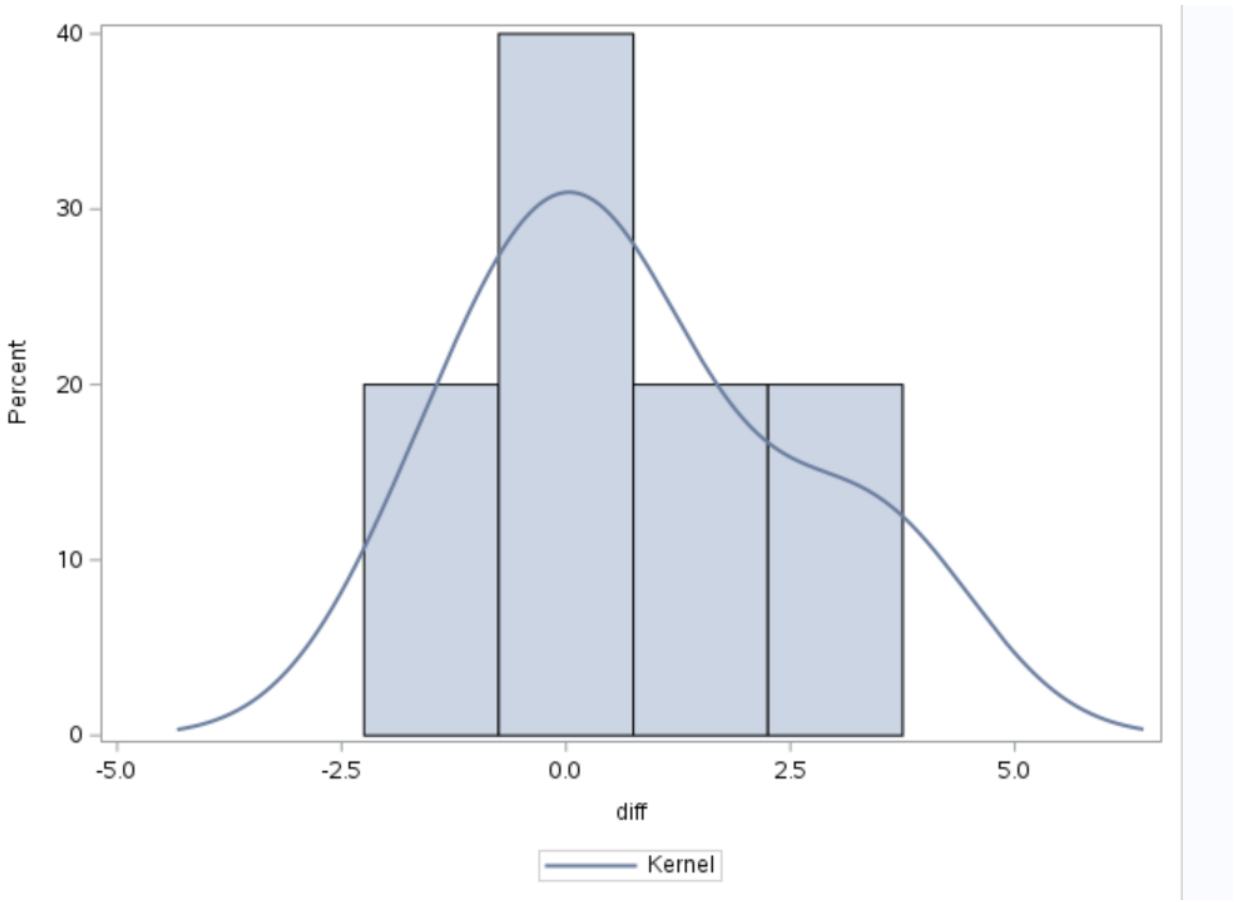


Figure 9: img 2

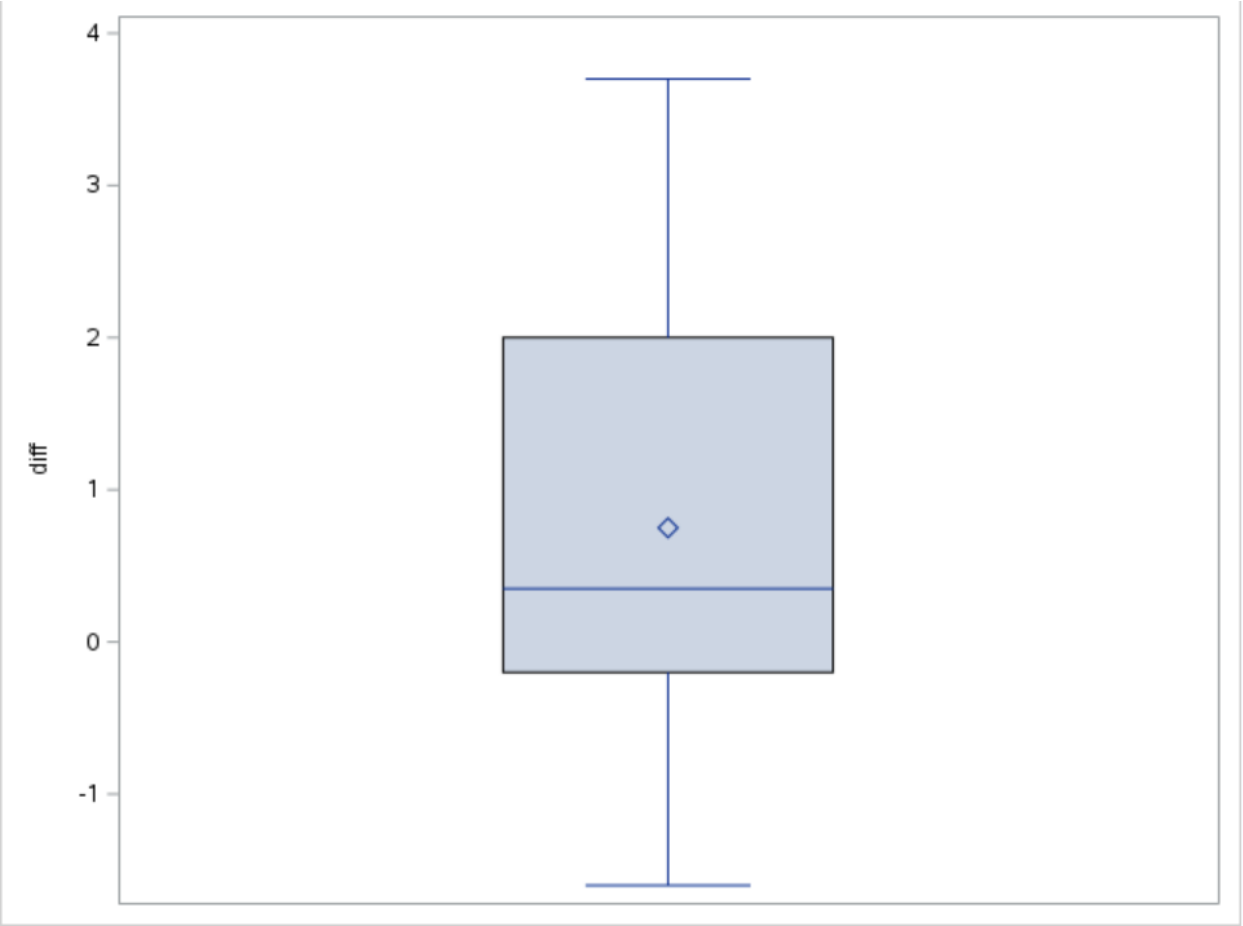


Figure 10: img 2

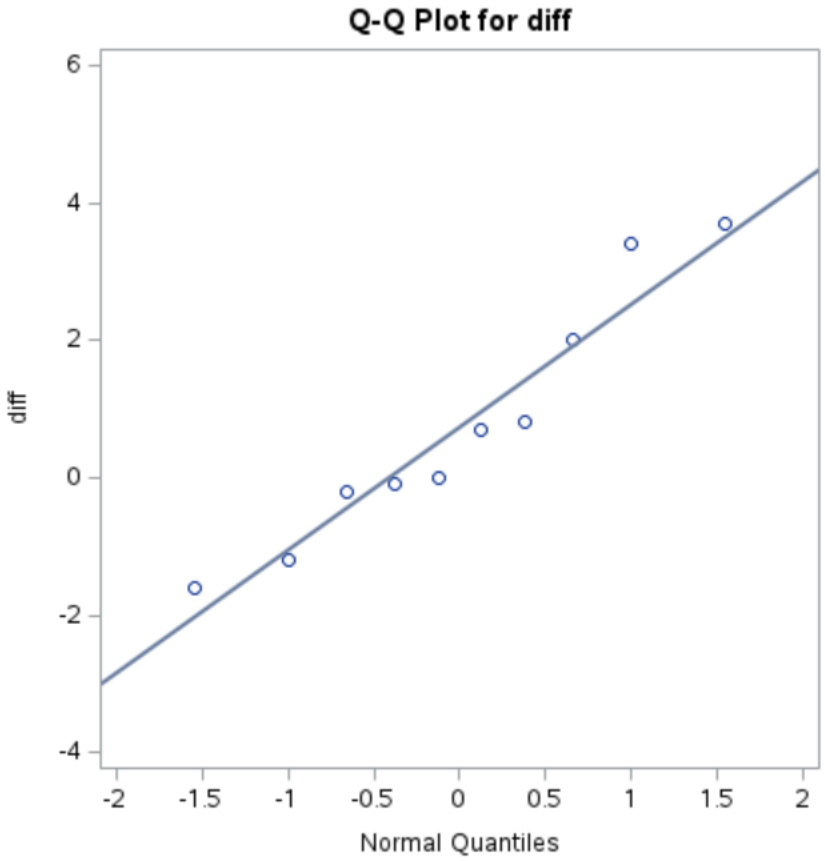
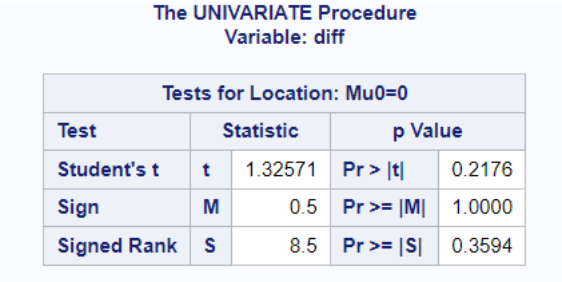


Figure 11: img 2

(f)

Apply the Wilcoxon signed rank test to these data in SAS. Report the (i) test statistic (sum of ranks for positive differences), (ii) p-value, and (iii) interpret the result in the context of the study.



The UNIVARIATE Procedure
Variable: diff

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	1.32571	Pr > |t|	0.2176
Sign	M	0.5	Pr >= |M|	1.0000
Signed Rank	S	8.5	Pr >= |S|	0.3594

Figure 12: img 2

Test Statistic: 8.5 p-value: 0.3594 Interpretation: We do not have statistically significant evidence to reject the null hypothesis that the distribution of sleep times for Placebo is different from the distribution of sleep times for the sleep-aid drug.

(g)

Apply the sign test to these data in SAS. Report the (i) test statistic (number of positive differences), (ii) p-value, and (iii) interpret the result in the context of the study

The UNIVARIATE Procedure
Variable: diff

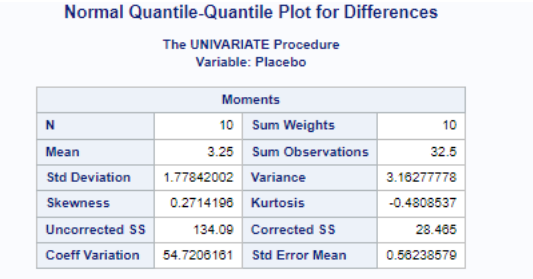
Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	1.32571	Pr > |t|	0.2176
Sign	M	0.5	Pr >= |M|	1.0000
Signed Rank	S	8.5	Pr >= |S|	0.3594

Figure 13: img 2

Test Statistic: 0.5 p-value: 1.000 Interpretation: We do not have statistically significant evidence to reject the null hypothesis. We do not have evidence that there are consistent differences between the sleep times of the Placebo and the sleep times of the sleep-aid drug. We do not have evidence of a direction of change, or, explicitly, that taking the sleep-aid drug would tend to be associated with a positive or negative change in the number of hours slept compared to the Placebo.

(h)

Compute how large (the number of patients) a study of this type would need to be so that the width of a 95% confidence interval for the difference in the mean responses to the drug and the placebo would be about 0.75 hours.

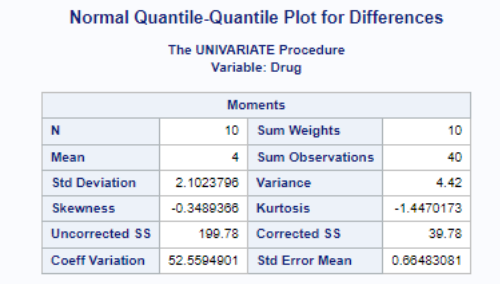


Normal Quantile-Quantile Plot for Differences

The UNIVARIATE Procedure
Variable: Placebo

Moments			
N	10	Sum Weights	10
Mean	3.25	Sum Observations	32.5
Std Deviation	1.77842002	Variance	3.16277778
Skewness	0.2714196	Kurtosis	-0.4808537
Uncorrected SS	134.09	Corrected SS	28.465
Coeff Variation	54.7206161	Std Error Mean	0.56238579

Figure 14: img



Normal Quantile-Quantile Plot for Differences

The UNIVARIATE Procedure
Variable: Drug

Moments			
N	10	Sum Weights	10
Mean	4	Sum Observations	40
Std Deviation	2.1023796	Variance	4.42
Skewness	-0.3489366	Kurtosis	-1.4470173
Uncorrected SS	199.78	Corrected SS	39.78
Coeff Variation	52.5594901	Std Error Mean	0.66483081

Figure 15: img

```
# qnorm()
# qt()
# equal n means
# S_p^2 = S_1^2 + S_2^2 / 2
sp1 <- 4.42
sp2 <- 3.1627778
spSq <- (sp1 + sp2) / 2
sp <- sqrt(spSq)
```

```
zval <- qnorm(p = 0.975, mean = 0, sd = 1)
n0 <- 8 * ((zval * sp)/0.75)^2
# 207.139
# use 208
tval <- qt(p = 0.975, df = 414)
n <- 8 * ((tval * sp)/0.75)^2
# qt(p = 0.975, )
n0
```

```
## [1] 207.139
```

```
n
```

```
## [1] 208.3555
```

209 Participants are needed so that the width of a 95% confidence interval for the difference in the mean responses to the drug and the placebo would be about 0.75 hours.