STAT 641 Fall 2021
Solutions for ASSIGNMENT 8

(P1) (8 pts.)

(S1) Matched pairs - both algorithms are applied to the same problems

(S2) Independent samples - There are two independent random samples, one from inner city schools and the other from suburban schools.

(S3) Matched pairs - Each of the 250 were observed under both stimuli.

(S4) Matched pairs - The two viruses were applied to the same leaf, one half to Virus 1 and the other half to Virus 2.

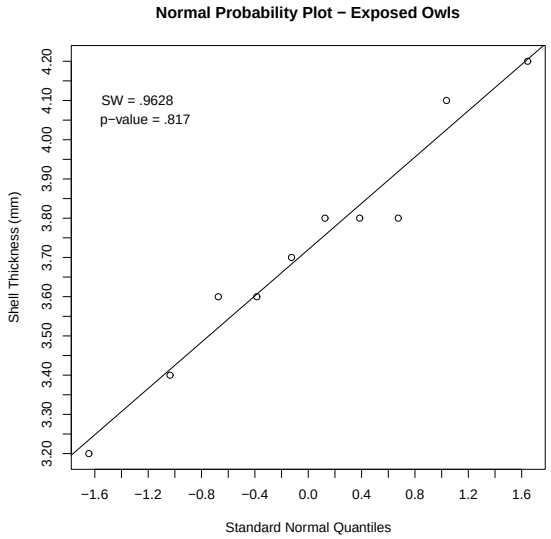
(P2) (8 pts.)

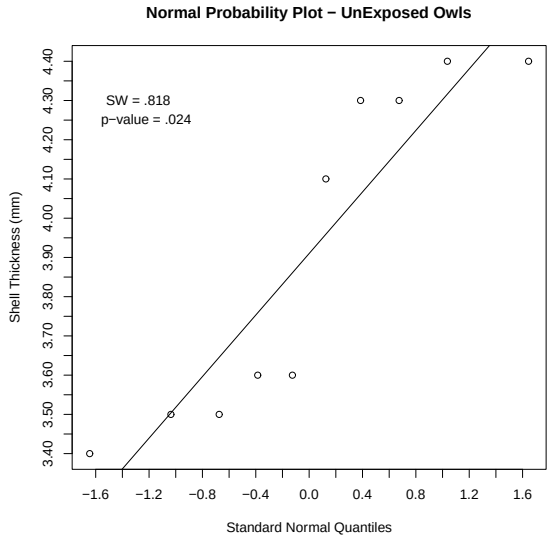
1. Test $H_o : \mu_{exp} \geq \mu_{unexp}$ vs $H_o : \mu_{exp} < \mu_{unexp}$. A separate variance t-test will be implemented because the p-value from the BFL test is .0265 (see part (E) for details) implies there is sufficient evidence to conclude that there is a difference in the two population's variances. From the data, the $C = \frac{(.2974)^2/10}{(.4228)^2/10} = .4948$ and $df = \frac{(.4948+1)^2(10-1)(10-1)}{(.4948)^2(10-1)+(10-1)} = 16.154$

$$t = \frac{3.72 - 3.91}{\sqrt{(.2974)^2/10 + (.4228)^2/10}} = -1.1623 > -1.745 = -t_{.05,16.154} \Rightarrow$$

$$\text{p-value } = P[t_{16.154} \leq -1.1623] = pt(-1.1623, 16.154) = .131$$

with 95% Upper Bound $(-\infty,\ .095)$ $\Rightarrow$ Fail to reject H_o and conclude that there is not significant evidence that the average shell thickness for eggs from PCB exposed owls is less than for unexposed owls. For the t-test to be valid both the exposed and unexposed shell thickness populations would need to have normal distributions. From the exposed data we have that the Shapiro-Wilk's test has p-value=.817 and the unexposed data has p-value =.024 . Thus, we would conclude that the data indicates that the exposed shell thickness have a normal distribution but the unexposed shell thickness do not have a normal distribution. This is confirmed by the following normal reference distribution plots:





Therefore, the p-value for the t-test may not be valid.

2. Because the distribution for the Unexposed shell thicknesses was not a normal distribution, the following calculations are not very reliable:

Let $\Delta = \frac{\mu_{exp}-\mu_{unexp}}{\sigma\sqrt{1/10+1/10}} = \frac{k\sigma}{\sigma\sqrt{2/10}} = k\sqrt{5}$, where k= 0, -.5, -1, -1.5, -2

$$\begin{aligned} P[\text{Type II error at } \mu_{exp}-\mu_{unexp}=k\sigma] &= P[\text{Fail to Reject } H_o \text{ at } \mu_{exp}-\mu_{unexp}=k\sigma] \\ &= P[t_{18,\Delta} > -1.734] \\ &= 1 - pt(-1.734, 18, \Delta] \end{aligned}$$

$\mu_{exp}-\mu_{unexp}$	0	$-.5\sigma$	-1.0σ	-1.5σ	-2.0σ
Δ	0	$-.5\sqrt{5}$	$-1.0\sqrt{5}$	$-1.5\sqrt{5}$	$-2.0\sqrt{5}$
P[Type II Error]	0	.7152	.3064	.0571	.0040

3. Using the normal based procedure with an estimated common variance of

$\hat{\sigma}^2 = \frac{(10-1)(.2974)^2+(10-1)(.4228)^2}{20-1} = (.3655)^2$, even though it is of questionable validity due to the conclusion of unequal variance and lack of normality,

$$m = \left(\frac{3+1}{3}\right)\frac{(\hat{\sigma})^2(Z_\alpha+Z_\beta)^2}{(\delta)^2} = \left(\frac{3+1}{3}\right)\frac{(.3655)^2(1.645+.84)^2}{(.3)^2} = 12.22 \Rightarrow m = 13 \text{ and } n = 3m = 39.$$

4. Using the Wilcoxon Rank Sum test (even though the distributions of the exposed and unexposed egg thicknesses are not in the same family of distributions), we obtain the following ranks:

Exposed	3.6	3.2	3.8	3.6	4.1	3.8	4.2	3.4	3.7	3.8
Ranks	7.5	1.0	12.0	7.5	14.5	12.0	16.0	2.5	10.0	12.0
UnExposed:	4.3	4.4	3.6	3.5	4.4	3.5	3.4	3.6	4.1	4.3
Ranks	17.5	19.5	7.5	4.5	19.5	4.5	2.5	7.5	14.5	17.5

The sum of the ranks for the Exposed owls is $W_1 = 95$, with

$p - value = pwilcox(W_1 - (10)(11)/2, 10, 10) = .2406$ which would imply that there is not sufficient evidence to conclude that the egg thickness for the exposed owls is shifted to the left of the egg thicknesses for the Unexposed owls.

Using the R-function: **wilcox.test(ex,uex,alternative="l",paired=FALSE)** we obtain

```
        Wilcoxon rank sum test with continuity correction
data:  ex and uex
W = 40, p-value = 0.2347
alternative hypothesis: true location shift is less than 0
Warning message:
In wilcox.test.default(ex, uex, alternative = "l", paired = FALSE) :
  cannot compute exact p-value with ties
```

5. There is not a valid test for testing $H_o : \sigma_{exp} \leq \sigma_{unexp}$ versus $H_1 : \sigma_{exp} > \sigma_{unexp}$ because the distribution of the shell thicknesses from the Unexposed owls have a non-normal distribution. We can test $H_o : \sigma_{exp} = \sigma_{unexp}$ versus $H_1 : \sigma_{exp} \neq \sigma_{unexp}$ using the Brown-Forsythe-Levene test (SAS) which yields a p-value of .0265 which would indicate there is sufficient evidence to conclude there is a difference in the variability of the two populations of egg thicknesses.

6. Based on the lack of normality in the data for the unexposed egg thicknesses, the t-test would not appear to be a valid procedure, especially considering the small sample sizes. Therefore, I would be more confident in using the results from the Wilcoxon Rank Sum test, although, the conditions for using this test are not valid either. However, the Wilcoxon Rank Sum test is much more robust against deviations from its conditions than is the t-test.

(P3) (8 pts.) Let $\tilde{\mu}_1$ and $\tilde{\mu}_2$ be the medians of Vitamin B and Placebo groups, respectively.

Test $H_0 : \tilde{\mu}_1 \leq \tilde{\mu}_2$ vs $H_1 : \tilde{\mu}_1 > \tilde{\mu}_2$.

Let $D = VitB - Placebo$. The values of D and the ranks of their magnitudes $|D_i|$ are given below:

Pair	1	2	3	4	5	6	7	8	9	10	11	12
$Y = X_1 - X_2$	6	8	9	5	-7	5	3	3	-12	3	0	1
Rank	7	9	10	5.5	8	5.5	3	3	11	3	*	1

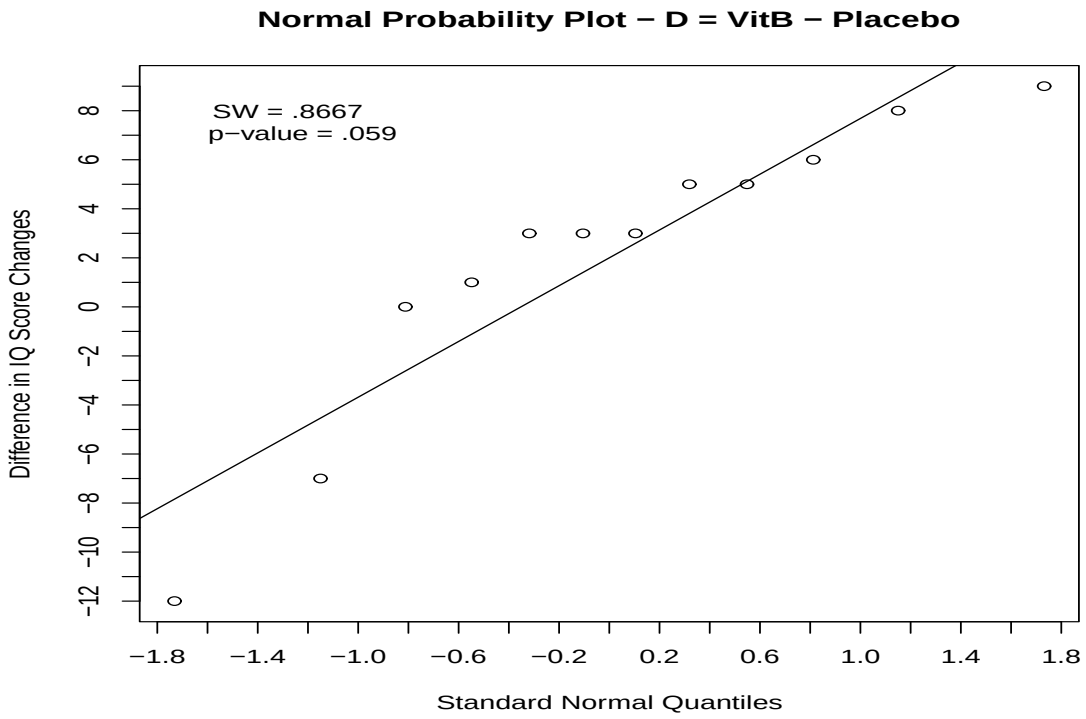
- **t-test:** $t = \frac{\bar{D}}{S_D/\sqrt{12}} = \frac{2}{6.0603/\sqrt{12}} = 1.143$ with $p-value = P[t_{12} \geq 1.143] = 1 - pt(1.143, 12) = .138$.
  Therefore, fail to reject H_o and conclude there is insufficient evidence that the median change in IQ score for the Vitamin B group is larger than the median change in IQ score for the Placebo group.

- **Signed Rank Test:** Let W_+ be sum of the ranks of positive differences.
  Note that the sample size is reduced to $n^* = 12 - 1$ due to one of the pairs having D=0.
  $W_+ = 47$ and $p - value = P[W_+ \geq 47] > P[W_+ \geq 48] = 0.103$ from Table A.10 in Textbook
  $p - value = P[W_+ \geq 47] = psignrank(46, 11, FALSE) = 0.120$ using R-function
  An approximate p–value is obtained using the Central Limit Theorem for W_+:
  $Z = \frac{W_+ - 0.5 - 33}{\sqrt{(11)(23)/2}} \sim N(0,1)$ approximately for large n

  $$p-value = P[W_+ \geq 47] \approx P\left[Z \geq \frac{47 - 0.5 - 33}{\sqrt{(11)(23)/2}}\right] = 1 - pnorm\left(\frac{47 - 0.5 - 33}{\sqrt{(11)(23)/2}}\right) = 1 - pnorm(1.20) = 0.115$$

  The following R code can be used to obtain the same results as above:

```
x = c(14,26,2,4,-5,14,3,-1,1,6,3,4)
y = c(8,18,-7,-1,2,9,0,-4,13,3,3,3)

t.test(x,y,alternative=c("greater"),paired=TRUE)

wilcox.test(x,y,alternative=c("greater"),paired=TRUE)

OUTPUT From R:
        Paired t-test

data:  V and P
t = 1.1432, df = 11, p-value = 0.1386
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -1.141826       Inf
sample estimates:
mean of the differences 2


OUTPUT From R:
       Wilcoxon signed rank test with continuity correction

data:  x and y
V = 47, p-value = 0.1144
alternative hypothesis: true location shift is greater than 0
```



- The Wilcoxon Signed Rank test provides a more reliable p-value than the paired t-test because the t-test requires that the differences have a normal distribution. From the normal reference distribution plot of the 12 differences, it would appear that a normal distribution is not a very good fit and the Shapiro-Wilk's test has p-value .059.

(P4) (8 pts.) (1.) Let p_1 be the probability that a Normal patient will have Low Excretions and p_2 be the probability that a Diabetic patient will have Low Excretions. We want to test $H_0 : p_1 = p_2$ vs $H_1 : p_1 \neq p_2$.

(2.) The $E'_{ij}s$ would be given by

$\hat{E}_{11} = (14)(12)/24 = 7,\ \hat{E}_{21} = (14)(12)/24 = 7,\ \hat{E}_{12} = (10)(12)/24 = 5,\ \hat{E}_{22} = (10)(12)/24 = 5$

We could use the Chi-square test because $E_{ij} \geq 5$ for all (i, j).

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}(O_{ij} - \hat{E}_{ij})^2/\hat{E}_{ij} = \frac{(10-7)^2}{7} + \frac{(4-7)^2}{7} + \frac{(2-5)^2}{5} + \frac{(8-5)^2}{5} = 6.171$$

With the two of the four expected values at 5, the Fisher Exact test will also be calculated:

$$p(x_0) = \frac{\binom{12}{x_0}\binom{12}{14-x_0}}{\binom{24}{14}} \quad \Rightarrow \quad p(10) = \frac{\binom{12}{10}\binom{12}{14-10}}{\binom{24}{14}} = dhyper(10, 12, 12, 14) = 0.0167.$$

For the Pearson Chi-squared Test: $p - value = P[\chi^2_1 > 6.171] = 1 = pchisq(6.171, 1) = .013$

For the Fisher Exact Test: p-value=$\sum_{x\in\Omega_0} p(x) = 0.0361$,

where $\Omega_0 = \{x : p(x) \leq p(x_0)\}$ and $p(x) = \frac{\binom{12}{x}\binom{12}{14-x}}{\binom{24}{14}} = dhyper(x, 12, 12, 14)$ for $2 \leq x \leq 12$.

x	2	3	4	5	6	7	8	9	10	11	12
p(x)	0.00003	0.00135	0.01666	0.08884	0.23321	0.31983	0.23321	0.08884	0.01666	0.00135	0.00003

Because of the small values of expected counts and the discrepancies between the two p-values, we will use the values from the Fisher Exact Test:

p-value $= .00003 + .00135 + .01666 + .01666 + .00135 + .00003 = 0.03608 < 0.05$,

Therefore, we reject H_0 and conclude that there is significant evidence (p-value=.036) of a difference in the urinary thromboglobulin excretion between normal and diabetic persons.

(P5) 8 pts.

1. $H_o : p_{1.} = p_{.1}$ vs $H_a : p_{1.} \neq p_{.1}$,

   where $p_{1.}, p_{.1}$ are the probabilities that drug 1 and drug 2 remained anesthetized, respectively.

   Because the data consists of the two responses from the same individual, the McNemar's test statistic is appropriate.

   The Pearson chi-square or Fisher Exact test would be inappropriate due to the correlation in the two responses.

2. $p - value = 2min(P[B \leq n_{12}], P[B \geq n_{12}]) = 2min(P[B \leq 13], P[B \geq 13]) = 2min(.9979, .0106) = .0212$,

   where B has a $Bin(m, .5) = Bin(3 + 13, .5) = Bin(16, .5)$ distribution.

   There is significant evidence (p-value=.0212) of a difference in the two drugs' probabilities that the membrane remained anesthetized.

(P6) 8 pts. Let p_1 = proportion of offspring that are Tall, cut-leaf,

p_2 = proportion of offspring that are Dwarf, cut-leaf,

p_3 = proportion of offspring that are Tall, potato-leaf,

p_4 = proportion of offspring that are Dwarf, potato-leaf.

Test the hypotheses

$H_0 : p_1 = \frac{9}{16}, \;\; p_2 = \frac{3}{16}, \;\; p_3 = \frac{3}{16}, \;\; p_4 = \frac{1}{16}$ vs H_1 : at least one of the p_i's differs from its specified value.

This is a multinomial distribution so we can use the chi-square goodness of fit test from Handout 9 to test the hypotheses as we did for testing the fit of a discrete distribution.

$\widehat{E}_1 = (1611)\left(\frac{9}{16}\right) = 906.1875, \;\; \widehat{E}_2 = \widehat{E}_3 = (1611)\left(\frac{3}{16}\right) = 302.0625, \;\; \widehat{E}_4 = (1611)\left(\frac{1}{16}\right) = 100.6875.$

i. T.S: $\chi^2 = \sum_{i=1}^{4} \frac{(O_i - \widehat{E}_i)^2}{\widehat{E}_i} = \frac{(926-906.1875)^2}{906.1875} + \frac{(293-302.0625)^2}{302.0625} + \frac{(288-302.0625)^2}{302.0625} + \frac{(104-100.6875)^2}{100.6875} = 1.4687$ with df=4-1=3

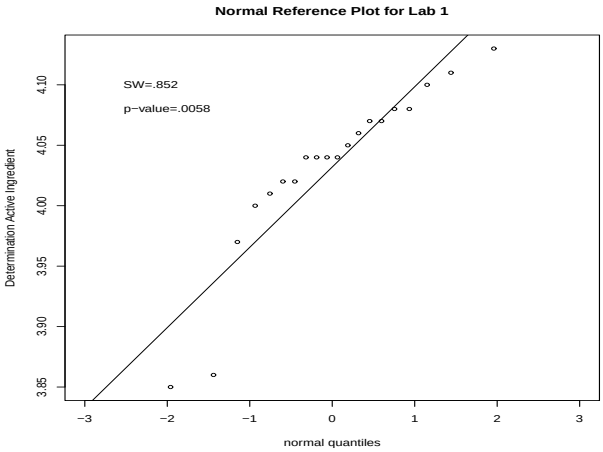
ii. p-value$=Pr(\chi^2_3 \geq 1.4687) = 1 - pchisq(1.4687, 3) = 0.6895 > 0.05 \quad \Rightarrow$ fail to reject H_0.

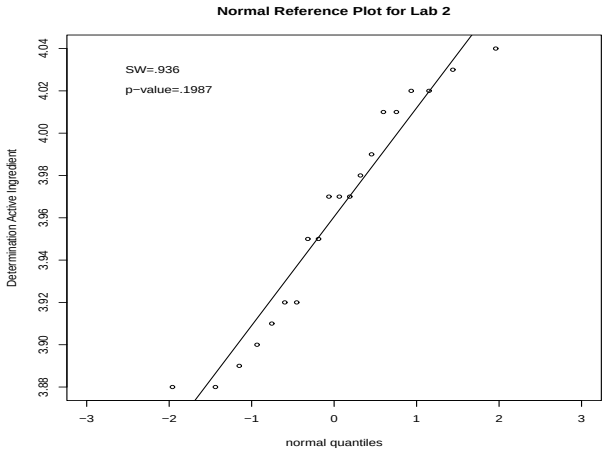
We conclude that there is not significant evidence (p-value=.6895) in the data that the tomato plants deviated from the current theory.

(P7) 14 pts.

1. Lab 1: plotted points are not close to the line and p-value=0.0058 from Shapiro-Wilk which implies that the normal distribution provides a poor fit to the data.

   Lab 2: plotted points are close to the line and p-value=0.1987 from Shapiro-Wilk which implies that the normal distribution provides a good fit to the data





2. $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$.

   From the data we have $\hat{\sigma}_1^2 = .005133$; $\hat{\sigma}_2^2 = .002752$

   From part 1. we have Lab 1 data has a non-normal distribution and therefore use BFL - test.

```
library(car)
y1 = c(4.13,4.07,4.04,4.07,4.05,4.04,4.02,4.06,4.10,3.86,
                  3.85,4.08,4.11,4.08,4.01,4.02,4.04,3.97,4.00,4.04)
y2 = c(3.88,4.02,4.01,4.01,4.04,3.99,4.03,3.97,3.98,4.02,
                 3.88,3.91,3.95,3.92,3.97,3.92,3.90,3.89,3.97,3.95)
y=c(y1,y2)
grp=as.factor(c(rep(1,20),rep(2,20)))
leveneTest(y, grp, center=median)
p-value=0.799
```

   The BFL test yields p-value=0.799 and hence we would fail to reject H_o.

   Thus, we conclude that there is not significant evidence that the two labs have different levels of variability.

- Note that the positive correlation for the data from Lab 2 found in part 3., would somewhat invalidate this conclusion.

3. Test the hypotheses: H_0 : Data not correlated vs H_1: Data is correlated
   i. $\hat{\rho}_1 = 0.260$ and $\hat{\rho}_2 = 0.313$ which indicates mild positive correlation in the readings from both labs
   ii. Because the data for Lab 1 was determined to be non-normally distributed, a Runs test will be used (see plots on next page)

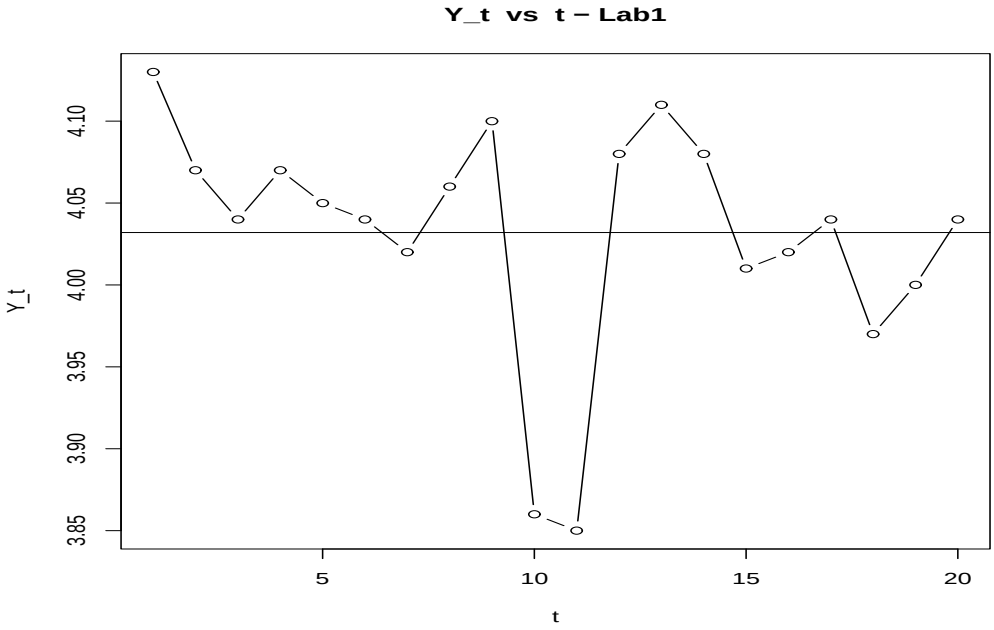
Runs test for Lab 1:
T.S: $r = \#$ of runs $= 9$.
$n_1 = 13,\ n_2 = 7 \quad \Rightarrow \quad$ From Table in Handout 13: $r_L = 5,\ r_U = 15$.
Because $r_L < r < r_U$, fail to reject H_0.

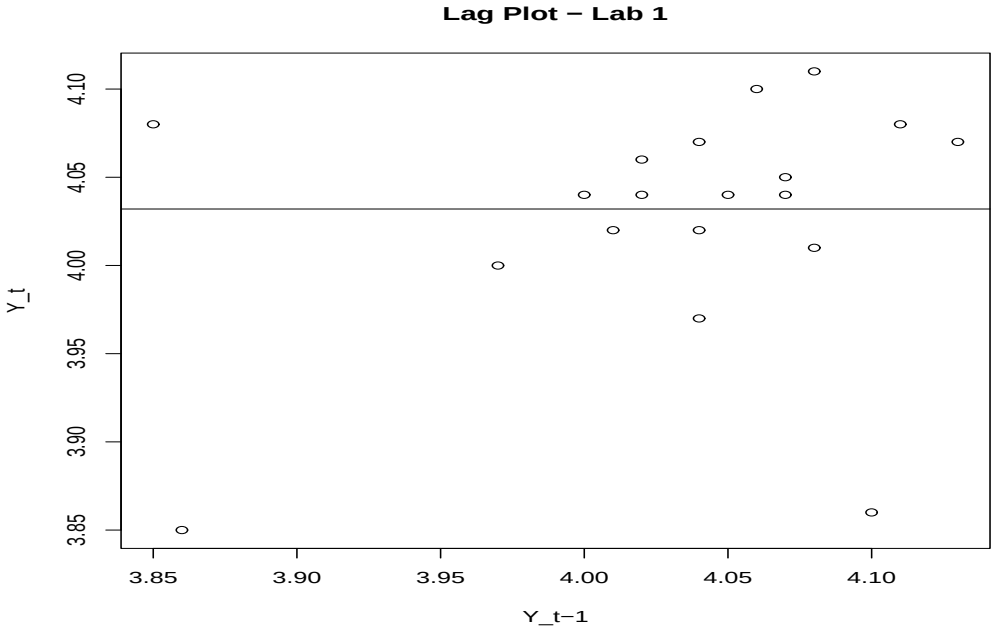
iii. Because there is a strong indication that the data is normally distributed, the von Neumann test will be used to evaluate correlation.
Reject H_o : Data is not correlated if $Q < Q_{P,.05} = 1.368$. From the data $\hat{\rho} = .3126$ and $Q = 1.314 < 1.368$. Conclude that there is sufficient evidence of correlation in the 20 observations from Lab 2.
Alternatively: Runs test for Lab 2:
$r = \#$ of runs $= 7$.
$n_1 = 11,\ n_2 = 9 \quad \Rightarrow \quad$ From Table in Handout 13: $r_L = 6,\ r_U = 16$.
Because $r_L < r < r_U$, fail to reject H_0.
When the data is from a normal distribution, the von Neumann test is more powerful that the Runs test and hence less likely to commit a Type II error.

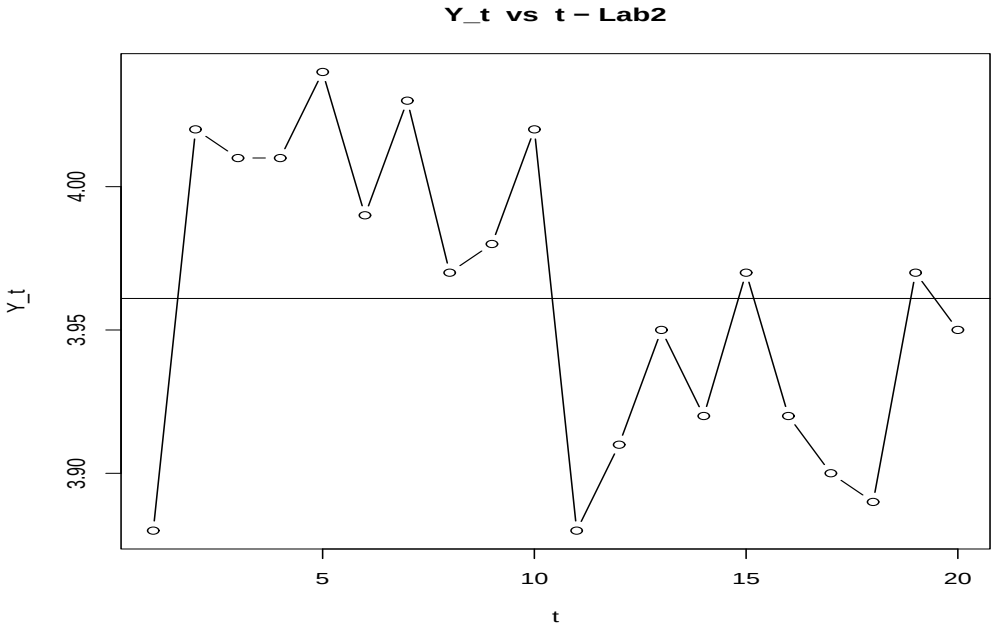
Thus we conclude that there is not significant evidence that the daily determinations within Lab 1 are correlated but there is mild positive correlation in the Lab 2 data.

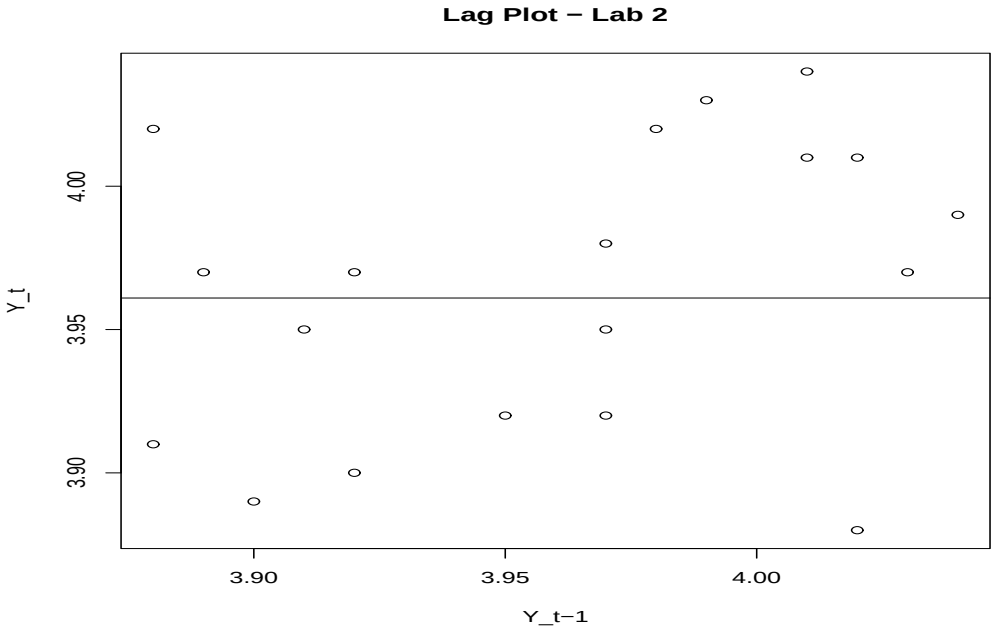
The graphs on the following pages display the plot of the data along with a lag plot. In the lag plot for Lab 1 there is no distinct pattern but a somewhat positive trend appears in the plot for Lab 2.

4. From part 1. we have that Lab 1 data is non-normal distributed and n_1 is relatively small. So, use Wilcoxon Rank sum test.
   i. Let $\tilde{\mu}_1$ and $\tilde{\mu}_2$ be the location parameters of the distributions for Lab 1 and Lab 2, respectively.
   Test: $H_0 : \tilde{\mu}_1 = \tilde{\mu}_2$ vs $H_1 : \tilde{\mu}_1 \neq \tilde{\mu}_2$.
   ii. T.S: Let W_1 and W_2 be sums of ranks from Lab 1 data and Lab 2 data, respectively, $W_1 = 541.5$ and $W_2 = 278.5$.
   iii. p-value$=2Pr(W_1 \geq w_{\max}) = 2Pr(W_1 \geq 541.5) = 2(1 - pwilcox(541 - 210, 20, 20) = .00020 < \alpha = 0.05 \quad \Rightarrow$ reject H_0.
   We conclude there is significant evidence that the two labs have the different average determinations.
   - Note that the positive correlation for the data from Lab 2 would somewhat invalidate this conclusion.


Y_t vs t – Lab1
Y_t
t
3.85
3.90
3.95
4.00
4.05
4.10
5
10
15
20


Lag Plot – Lab 1
Y_t
Y_t−1
3.85
3.90
3.95
4.00
4.05
4.10





5. Because the data from Lab 1 appears non-normal and the sample size, $n_1 = 20$ is small, we could report a distribution-free confidence interval on the medians instead of the mean:

   From Table VII.3 on page 32 in Handout 11, we have that with k=6, a 95.9% C.I. on the median determination is $(X_{(6)},\ X_{(15)})$

   For Lab 1, a 95.9% C.I. on the median determination is (4.02, 4.07)

   Alternatively, we could perform a studentized bootstrap to get an approximate C.I. on the mean. Using the provided code, the resulting interval is (3.98, 4.06)

   For Lab 2, a 95% C.I. based on studentized bootstrap is (3.93, 3.98)

The distribution-free 95.9% C.I. on the median determination for Lab 2 is $(X_{(6)},\ X_{(15)}) = (3.92,\ 4.01)$

(P8) 8 pts. Let p_{ij} be the probability that a randomly selected child has the ith level of Tonsil Size and jth level of Carrier Status.

Test for Independence between Tonsil Size and Carrier Status:

$H_0 : p_{ij} = p_{i\cdot}p_{\cdot j}$ for all pairs (i,j) vs $H_1 : p_{ij} \neq p_{i\cdot}p_{\cdot j}$ for some (i,j)

$$\widehat{E}_{11} = \frac{516)(72)}{1398} = 26.58, \quad \widehat{E}_{12} = \frac{516)(1326)}{1398} = 489.42, \quad \widehat{E}_{21} = \frac{589)(72)}{1398} = 30.33,$$
$$\widehat{E}_{22} = \frac{589)(1326)}{1398} = 558.67, \quad \widehat{E}_{31} = \frac{293)(72)}{1398} = 15.09, \quad \widehat{E}_{32} = \frac{293)(1326)}{1398} = 277.91$$

$\chi^2 = \sum_{i=1}^{3}\sum_{j=1}^{2} \frac{(O_{ij}-\widehat{E}_{ij})^2}{\widehat{E}_{ij}} = 7.885$ with $df = (3-1)(2-1) = 2 \quad \Rightarrow$

p-value=$Pr(\chi^2_2 \geq 7.885) = 1 - pchisq(7.885, 2) = 0.0194 < 0.05 \quad \Rightarrow$ reject H_0.

We thus conclude that there is significant evidence (p-value=.0194) that Tonsil Size and Carrier Status are associated.

(P9) (16 points) Multiple Choice.

MC1 **C** because of the strong positive correlation in the data.

MC2 **D** because $t = \frac{\bar{X}-53}{S/\sqrt{200}} > \frac{\bar{X}-53}{\widehat{SE}(\bar{X})} = t^* \ \Rightarrow\ P[t_{199} \geq t] < P[t_{199} \geq t^*]$

MC3 **C** See HO 12 discussion on the robustness of the test of variances

MC4 **B** Testing $H_o : \sigma \leq 2.5$ vs $H_1 : \sigma > 2.5$. Size of the test is the height of the curve when $\sigma = 2.5$

MC5 **A** P[Type II error at $\sigma = 3.5$] $= 1 - \gamma(3.5) = 1 - .79 = .21$

MC6 **C** With $n = 40,\ \alpha = .01$, Reject H_o when $\frac{(n-1)S^2}{(2.5)^2} \geq \chi^2_{39,.01} = qchisq(.99, 39) = 62.43$

P[Type II error at $\sigma = 2.775$] $= P\left[\frac{(n-1)S^2}{(2.5)^2} < \chi^2_{39,.01} \text{ when } \sigma = 2.775\right] =$

$P\left[\frac{(n-1)S^2}{(2.755)^2} < \frac{(2.5)^2}{(2.775)^2}\chi^2_{39,.01} \text{ when } \sigma = 2.775\right] =$

$P[\chi^2_{39} < \frac{(2.5)^2}{(2.775)^2}62.43] = P[\chi^2_{39} < 50.66] = pchisq(50.66, 39) = .90$

MC7 **B** Using Table A11 from Handout 12, the value of n is 27.
Using the R function:
power.t.test(n=,delta=.2,sd=.4,sig.level=.05,power=.8,type=c("one.sample"),alternative=c("one.sided")), n = 27
Using the formula for the Z-test, $n = \frac{(.4)^2[1.645+.84]^2}{(3-2.8)^2} = 24.7$, thus this simple formula underestimates the necessary sample size. The power using n=25, is .783, whereas, using n=27, the power is .812.

MC8 **B or D** Pairing increases the power of the paired t-test over the independent samples t-test only when there is a strong positive correlation between the responses within each of the n pairs of experimental units.

```
####
#### (2)
####

x_exp <- c(3.6, 3.2, 3.8, 3.6, 4.1, 3.8, 4.2, 3.4, 3.7, 3.8)
x_nexp <- c(4.3, 4.4, 3.6, 3.5, 4.4, 3.5, 3.4, 3.6, 4.1, 4.3)
n <- length(x_exp)

##
## (1)
##

## Are the data Normally distributed? Not the not-exposed data apparently.
shapiro.test(x_exp)
shapiro.test(x_nexp)

## Are the variances equal? BFL test p-value = 0.0265.
library(car)

x_all <- c(x_exp, x_nexp)
grp <- factor(rep(1:2, each = n))

leveneTest(x_all, grp)

## Go ahead and use an unequal-variance t-test.
t.test(x_exp, x_nexp, alt = "less", var.equal = FALSE)

## Compute approximate df for unequal-variance t-test.
s_1 <- sd(x_exp)
s_2 <- sd(x_nexp)
CC <- (s_1 ^ 2 / n) / (s_2 ^ 2 / n)
nu <- ((CC + 1) ^ 2 * (n - 1) * (n - 1)) / (CC ^ 2 * (n - 1) + n - 1)

##
## (2)
##

## Probability of Type II error for pooled t-test
s_p <- sqrt(((n - 1) * var(x_exp) + (n - 1) * var(x_nexp)) / (n + n - 2))
k <- c(0, -0.5, -1, -1.5, -2)
Delta <- k * sqrt(5)

qt(0.05, 18)

## The probability of a Type II error when Delta = 0 is actually 0
for(i in 1:length(k))
  print(1 - pt(-1.734064, 18, Delta[i]))

##
## (3)
##

delta <- -0.3
m <- ((3 + 1) / 3) * (s_p * (qnorm(0.95) + qnorm(0.8)) / delta) ^ 2
```

```
##
## (4)
##

x_all <- c(x_exp, x_nexp)
rk_all <- rank(x_all)
W_1 <- sum(rk_all[1:n])
p_val <- pwilcox(W_1 - n * (n + 1) / 2, n, n)

wilcox.test(x_exp, x_nexp, alt = "less", paired = FALSE)

##
## (5) See above use of 'leveneTest'
##

####
#### (3)
####

x_vit <- c(14, 26, 2, 4, -5, 14, 3, -1, 1, 6, 3, 4)
x_pla <- c(8, 18, -7, -1, 2, 9, 0, -4, 13, 3, 3, 3)
d <- x_vit - x_pla
n <- length(x_vit)

## Paired differences normal?
qqnorm(d); qqline(d)
shapiro.test(d)

## Do both paired T-test and Wilcoxon signed rank test.
t.test(d, alt = "greater")
wilcox.test(x_vit, x_pla, alt = "greater", paired = TRUE)

####
#### (4)
####

## Chi-square test
N <- 24
R_1 <- 14
R_2 <- 10
n_1 <- 12
n_2 <- 12
p_0 <- R_1 / N

O_11 <- 10
O_12 <- 2
O_21 <- 4
O_22 <- 8
E_11 <- n_1 * p_0
E_12 <- n_1 * (1 - p_0)
E_21 <- n_2 * p_0
E_22 <- n_2 * (1 - p_0)
OO <- c(O_11, O_12, O_21, O_22)
```

```
EE <- c(E_11, E_12, E_21, E_22)

chi_sq <- sum((OO - EE) ^ 2 / EE)
p_val <- 1 - pchisq(chi_sq, 1)

## Fisher's Exact test
p_x <- dhyper(0:12, 12, 12, 14)
p_val <- sum(p_x[p_x <= p_x[11]])

M <- matrix(c(O_11, O_21, O_12, O_22), nrow = 2)
fisher.test(M, alternative = "two.sided")

####
#### (5)
####

## McNemar's test
m <- 16
n_12 <- 13
p_val <- 2 * min(1 - pbinom(n_12 - 1, m, 0.5), pbinom(n_12, m, 0.5))

####
#### (6)
####

## Chi-square GOF test for completely specified distribution
OO <- c(926, 293, 288, 104)
N <- sum(OO)
p_0 <- c(9 / 16, 3 / 16, 3 / 16, 1 / 16)
EE <- N * p_0

chi_sq <- sum((OO - EE) ^ 2 / EE)
p_val <- 1 - pchisq(chi_sq, 3)

####
#### (7)
####

x_1 <- c(4.13, 4.07, 4.04, 4.07, 4.05, 4.04, 4.02, 4.06, 4.10, 3.86, 3.85, 4.08, 4.11,
  4.08, 4.01, 4.02, 4.04, 3.97, 4.00, 4.04)
x_2 <- c(3.88, 4.02, 4.01, 4.01, 4.04, 3.99, 4.03, 3.97, 3.98, 4.02, 3.88, 3.91, 3.95,
  3.92, 3.97, 3.92, 3.90, 3.89, 3.97, 3.95)
n <- length(x_1)

##
## (1)
##

## Reference distribution plots
x_1_sort <- sort(x_1)
x_2_sort <- sort(x_2)
u <- (1:n - 0.5) / n
Q <- qnorm(u)
```

```
plot(Q, x_1_sort, xlab = "Normal Quantiles", ylab = "Sample Quantiles", main = "Lab 1")
abline(lm(x_1_sort ~ Q))

plot(Q, x_2_sort, xlab = "Normal Quantiles", ylab = "Sample Quantiles", main = "Lab 2")
abline(lm(x_2_sort ~ Q))

## Shapiro-Wilks tests
shapiro.test(x_1)
shapiro.test(x_2)

##
## (2)
##

require(car)

## BFL test
x <- c(x_1, x_2)
grp <- factor(rep(1:2, each = n))

leveneTest(x, grp, center = median)

##
## (3)
##

## Sample autocorrelation estimates
x_bar_1 <- mean(x_1)
x_bar_2 <- mean(x_2)
rho_1 <- sum((x_1[2:n] - x_bar_1) * (x_1[1:(n - 1)] - x_bar_1)) /
  sum((x_1 - x_bar_1) ^ 2)
rho_2 <- sum((x_2[2:n] - x_bar_2) * (x_2[1:(n - 1)] - x_bar_2)) /
  sum((x_2 - x_bar_2) ^ 2)

## Runs test for Lab 1
x_1_c <- x_1 - x_bar_1
n_1 <- sum(x_1_c > 0)
n_2 <- sum(x_1_c < 0)

numb_runs <- 1
for(j in 2:n) {
  if(sign(x_1_c[j]) != sign(x_1_c[j - 1]))
    numb_runs <- numb_runs + 1
}

## von Neumann test for Lab 2
Q <- (1 / (n - 1)) * sum((x_2[2:n] - x_2[1:(n - 1)]) ^ 2) /
  ((1 / n) * sum((x_2 - x_bar_2) ^ 2))

## Time series plots
x_ts_1 <- ts(x_1, start = 1, frequency = 1)
x_ts_2 <- ts(x_2, start = 1, frequency = 1)
plot.ts(x_ts_1, type = "b", xlab = "t", ylab = "x_t", main = "Lab 1 Time Series Plot")
abline(x_bar_1, 0)
```

```
plot.ts(x_ts_2, type = "b", xlab = "t", ylab = "x_t", main = "Lab 2 Time Series Plot")
abline(x_bar_2, 0)

## Lag plots
plot(x_1[1:(n - 1)], x_1[2:n], xlab = "x_t-1", ylab = "x_t", main = "Lab 1 Lag Plot")
abline(x_bar_1, 0)
plot(x_2[1:(n - 1)], x_2[2:n], xlab = "x_t-1", ylab = "x_t", main = "Lab 2 Lag Plot")
abline(x_bar_2, 0)

##
## (4)
##

## Wilcoxon Rank Sum test
x_all <- c(x_1, x_2)
x_all_r <- rank(x_all)
x_1_r <- x_all_r[1:n]
x_2_r <- x_all_r[(n + 1):(2 * n)]

W_1 <- sum(x_1_r)
W_2 <- sum(x_2_r)
p_val <- 2 * min(1 - pwilcox(W_1 - n * (n + 1) / 2 - 1, n, n),
  pwilcox(W_1 - n * (n + 1) / 2, n, n))

wilcox.test(x_1, x_2, alternative = "t", paired = FALSE)

##
## (5)
##

## Since Lab 1 did not appear to be Normally distributed, could do distribution-free CI
## on median. Alternatively, could do studentized bootstrap CI on mean.
B <- 9999
theta_1 <- mean(x_1)
V_1 <- var(x_1) / n

Z_star <- numeric(B)
for(b in 1:B) {
  x_b <- sample(x_1, replace = TRUE)
  theta_b <- mean(x_b)
  V_b <- var(x_b) / n
  Z_star[b] <- (theta_b - theta_1) / sqrt(V_b)
}
Z_star_o <- sort(Z_star)
theta_1 - sqrt(V_1) * Z_star_o[9750]
theta_1 - sqrt(V_1) * Z_star_o[250]

## Repeat studentized bootstrap on Lab 2.
theta_2 <- mean(x_2)
V_2 <- var(x_2) / n

Z_star <- numeric(B)
for(b in 1:B) {
  x_b <- sample(x_2, replace = TRUE)
```

```
  theta_b <- mean(x_b)
  V_b <- var(x_b) / n
  Z_star[b] <- (theta_b - theta_2) / sqrt(V_b)
}
Z_star_o <- sort(Z_star)
theta_2 - sqrt(V_2) * Z_star_o[9750]
theta_2 - sqrt(V_2) * Z_star_o[250]

####
#### (8)
####

## Chi square test of independence
N <- 1398

O_11 <- 19
O_12 <- 497
O_21 <- 29
O_22 <- 560
O_31 <- 24
O_32 <- 269

n_1_dot <- 516
n_2_dot <- 589
n_3_dot <- 293
n_dot_1 <- 72
n_dot_2 <- 1326
p_i_dot <- c(n_1_dot, n_2_dot, n_3_dot) / N
p_dot_i <- c(n_dot_1, n_dot_2) / N

E_11 <- N * p_i_dot[1] * p_dot_i[1]
E_12 <- N * p_i_dot[1] * p_dot_i[2]
E_21 <- N * p_i_dot[2] * p_dot_i[1]
E_22 <- N * p_i_dot[2] * p_dot_i[2]
E_31 <- N * p_i_dot[3] * p_dot_i[1]
E_32 <- N * p_i_dot[3] * p_dot_i[2]

OO <- c(O_11, O_12, O_21, O_22, O_31, O_32)
EE <- c(E_11, E_12, E_21, E_22, E_31, E_32)
chi_sq <- sum((OO - EE) ^ 2 / EE)
p_val <- 1 - pchisq(chi_sq, 2)

## Using chisq.test()
M <- matrix(c(O_11, O_21, O_31, O_12, O_22, O_32), nrow = 3)
chisq.test(M)
```

STATISTICS 641 - ASSIGNMENT 8

DUE DATE: 11:59pm, WEDNESDAY, DECEMBER 8, 2021

Name _____

Email Address _____

Make this cover sheet the first page of your Solutions.

STATISTICS 641 - ASSIGNMENT 8

- **Due 11:59pm Wednesday, December 8, 2021**
- Read Handout 13
- Supplemental Reading from Devore book: Chapters 9, 10, 14, & 15

(P1) (8 points) In each of the following studies, state whether the study uses an independent samples or a matched pairs design:

(S1) In an evaluation of the efficiency of algorithms, two algorithms are evaluated in terms of CPU times required to complete the same six test problems.

(S2) A survey is conducted of 16 year old students from inner city public schools and suburban public schools to compare the proportion who had experimented with marijuana.

(S3) A psychologist designs a study to assess whether a visual or audio stimulus produces a more rapid response. A group of 250 undergraduates are randomly assigned to the order in which they are exposed to the two stimuli, audio then visual or vice versa. The response times to the stimuli are then recorded.

(S4) The effect of two types of viruses on tobacco leaves was studied by rubbing a preparation containing one of the viruses onto a different half of each of 8 tobacco leavers. The number of lesions counted on the two halves of these leavers were recorded.

(P2) (8 points) An experiment is run to study the effects of PCB, an industrial contaminant, on the reproductive ability of owls. The shell thickness (mm) of eggs produced by 10 owls exposed to PCB are compared to the shell thickness of eggs produced by 10 owls which did not have PCB exposures.

Owl	1	2	3	4	5	6	7	8	9	10
PCB-Exposed:	3.6	3.2	3.8	3.6	4.1	3.8	4.2	3.4	3.7	3.8
UnExposed:	4.3	4.4	3.6	3.5	4.4	3.5	3.4	3.6	4.1	4.3

1. Is there significant($\alpha = .05$) evidence that the PCB exposed owls have thinner egg shells than those of the unexposed owls? Use a t-test in reaching your conclusion and report the p-value.
2. Compute the chance that your test committed a Type II error for the following values of $\theta = \mu_{exposed} - \mu_{unexposed} = 0, -.5\sigma, -\sigma, -1.5\sigma, -2\sigma$. Base this calculation on the pooled t-test.
3. In designing a new study, the researchers want to determine the necessary sample sizes for exposed and unexposed owls such that an $\alpha = .05$ test will have power of at least 80% to detect a shell thickness difference of more than 0.3 mm. The researchers want to examine three times as many exposed owls as unexposed owls, that is, m=3n.
4. Is there significant($\alpha = .05$) evidence that the PCB exposed owls have thinner egg shells than those of the unexposed owls? Use a Wilcoxon test in reaching your conclusion and report the p-value.
5. Is there significant($\alpha = .05$) evidence that the PCB exposed owls have greater variability in egg shell thickness than those of the unexposed owls? Report the p-value of your test.
6. Which test, t-test or Wilcoxon, is more appropriate for testing the difference in egg shell thickness?

(P3) (8 points) In a study of the effect of vitamin B on learning, 12 pairs of children were matched on IQ, age, size, and general health. Within each pair, one child was randomly selected to receive a vitamin B table every day and the other child received a placebo tablet. The following table shows the change in IQ score over the six months of the study.

Pair	1	2	3	4	5	6	7	8	9	10	11	12
Vitamin B	14	26	2	4	-5	14	3	-1	1	6	3	4
Placebo	8	18	-7	-1	2	9	0	-4	13	3	3	3

Is there substantial evidence that a six months treatment with vitamin B increased IQ score? Use $\alpha = .05$ in applying both the t-test and the Wilcoxon signed rank test for these hypotheses. Which test produces the most reliable conclusion?

(P4) (8 points) A study evaluated the urinary-thromboglobulin excretion in 12 normal and 12 diabetic patients. The excretions were summarized with a value of 20 or less labeled as "Low" and values above 20 as "High".

	Excretion	
	Low	High
Normal	10	2
Diabetic	4	8

1. Set up hypotheses to assess whether there is substantial evidence of a difference in the urinary-thromboglobulin excretion between normal and diabetic patients.
2. At the $\alpha = .05$ level what can you conclude? Report a p-value for your test.

(P5) (8 points) A study was conducted to compare two topical anesthetic drugs for use in dentistry. The two drugs were applied on the oral mucous membrane of the two sides of each patient's mouth and after a fixed period of time it was recorded whether or not the membrane remained anesthetized. Data from the 45 patients is recorded below:

		Drug 2 Response	
		Anesthetized	Not Anesthetized
Drug 1	Anesthetized	15	13
Response	Not Anesthetized	3	14

1. Set up hypotheses to assess whether there is substantial evidence of a difference between the two drugs.
2. At the $\alpha = .05$ level what can you conclude? Report a p-value for your test.

(P6) (8 points) A genetics experiment on the characteristics of tomato plants provided the following data on the number of offspring expression four phenotypes.

Phenotype	Tall, cut-leaf	Dwarf, cut-leaf	Tall, potato-leaf	Dwarf, potato-leaf	Total
Frequency	926	293	288	104	1611

The researcher wants to determine if there is substantial evidence that the tomato plants deviate from the current theory that the four phenotypes will appear in the proportion 9:3:3:1. Use $\alpha = .01$.

(P7) (14 points) A company is attempting to automate the determination of the amount of the active ingredient, chlorphaniramine maleate, in the tablets it produces. Two labs were asked to make 20 determinations on a composite sample which had a nominal dosage level of 4 milligrams. The purpose of the experiment was to study the consistency between labs and the variability of the determination procedure within labs. The data is given in the following table.

	Day									
	1	2	3	4	5	6	7	8	9	10
Lab 1	4.13	4.07	4.04	4.07	4.05	4.04	4.02	4.06	4.10	3.86
Lab 2	3.88	4.02	4.01	4.01	4.04	3.99	4.03	3.97	3.98	4.02

	Day									
	11	12	13	14	15	16	17	18	19	20
Lab 1	3.85	4.08	4.11	4.08	4.01	4.02	4.04	3.97	4.00	4.04
Lab 2	3.88	3.91	3.95	3.92	3.97	3.92	3.90	3.89	3.97	3.95

1. Do the readings from the labs appear to have a normal distribution? Justify your answer.
2. Do the readings from the two labs appear to have the same level of variability? Justify your answer.
3. Do the daily determinations within each lab appear to be correlated? Justify your answer.
4. Do the readings from the two labs appear to have different average determinations? Justify your answer.
5. Provide 95% confidence intervals on the average determinations for both labs.

(P8) (8 points) A study was conducted to investigate whether there is a relationship between tonsil size and carriers of a particular bacterium, *Streptococcus pyrogenes*. The following table contains the results from 1398 children.

Tonsil	Carrier Status		Row
Size	Carrier	Noncarrier	Total
Normal	19	497	516
Large	29	560	589
Very Large	24	269	293
Column Total	72	1326	1398

Is there significant evidence that tonsil size and carrier status are associated? Use $\alpha = .05$.

(P9) (16 points) Multiple Choice Questions.

CIRCLE ONE of the following letters **(A, B, C, D, or E)** corresponding to the **BEST** answer.

Use the following information for MC1 and MC2.

A process engineer samples a continuous flow of the company's product 200 times per day and obtains the following pH levels in the product : $X_1, \cdots, X_{200}$. He determines that the daily pH levels are related by $X_t = \theta + \rho X_{t-1} + e_t$, where the e_ts have independent $N(0, \sigma_e^2)$ distributions and $\rho > .9$.

MC1 The engineer constructs a nominal 95% confidence interval for the average daily pH level, μ, using the formula $\bar{X} \pm t_{.025,199}(s/\sqrt{200})$, where $\bar{X}$ and s are the sample mean and standard deviation for a given days pH levels. The true coverage probability of this confidence interval

A. is 0.95.

B. is very close to 0.95.

C. is much less than 0.95.

D. is much greater than 0.95.

E. may be greater than 0.95 or less than 0.95 depending on the distribution of the X_t's.

MC2 Refer to MC1. The nominal pH level of the product is 5.3. The process engineer wants to test if the pH on a given day is different from 5.3. He uses $t = \frac{\bar{X}-5.3}{s/\sqrt{200}}$ as his test statistic. Next, he uses the t-distribution with df=199 to compute the p-value of the observed data. The computed p-value will be

A. correct because the sample size is large.

B. very close to the correct p-value because the sample size is large.

C. much larger than the correct p-value.

D. much smaller than the correct p-value.

E. may be greater or less than the correct value depending on the size of *sigma*.

MC3 A psychologist is investigating the IQ level of young children who have been in a head start program. She wants to determine if the variation in IQ scores for the population of head start students is smaller than the variation in the general population of children under the age of 6 which has a variation of $\sigma = 10.2$. She also informs you that the distribution of IQ scores is highly right skewed. Suppose she uses the test: reject H_o is $\frac{(n-1)S^2}{(10.2)^2} < \chi^2_{.95,n-1}$, where S is the standard deviation from a random sample of n head start students, to test whether σ is less than 10.2 with an α value of 0.05.

A. the actual level of significance will be very close to 0.05.

B. the actual level of significance will be less than 0.05.

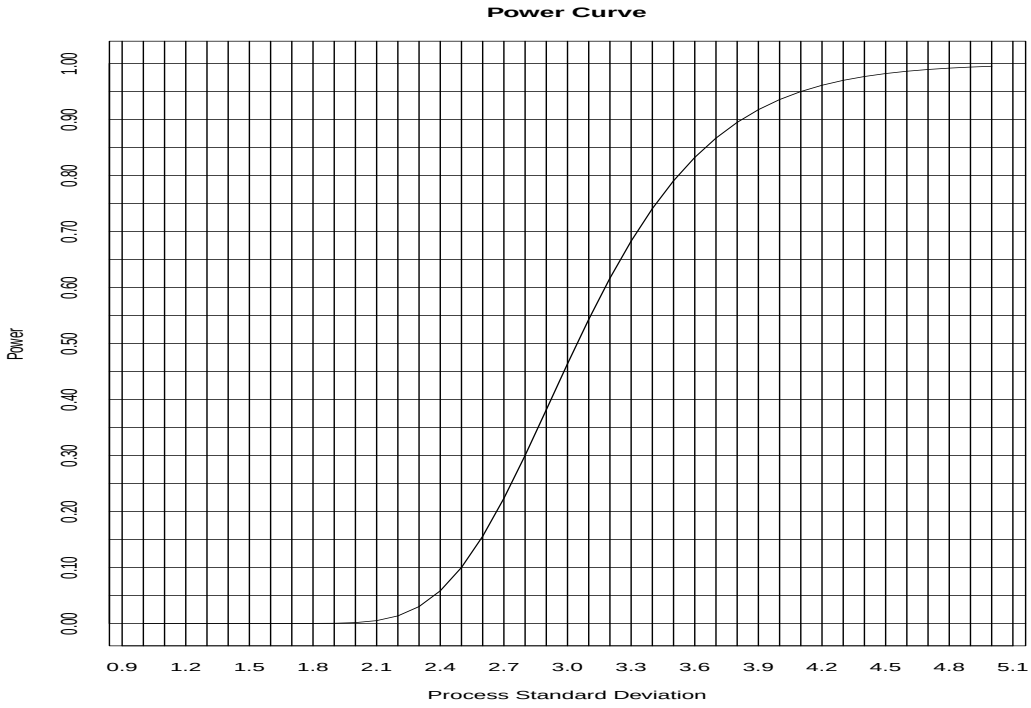
C. the actual level of significance will be greater than 0.05.

D. the actual level of significance will be exactly 0.05.

E. it is impossible to determine the effect of skewness on the actual level of significance.

The following discussion will supply the information for Questions MC4, MC5, and MC6.

A company that manufacturers silicon wafers for computer chips is concerned with both the mean thickness of the chips and the fluctuation in the thickness of the chips. In order to monitor the thickness, a random sample of n chips is selected every hour and the thickness is measured on each of the chips. The process is considered to be in control provided the process mean, μ, is 200 mm and the process standard deviation, σ is less than or equal to 2.5 mm. The company's process engineer develops a test to evaluate whether the process standard deviation is greater than 2.5 mm. She plots the power curve of the test in order to evaluate its performance. The curve is given here:



MC4 What is the maximum probability of a Type I error when using the test whose power curve is depicted above?

A. .05

B. .10

C. .15

D. .90

E. .95

MC5 What is the probability of a Type II error of the test in MC4 if $\sigma = 3.5$?

A. .21

B. .31

C. .69

D. .79

E. cannot be determined from the power curve

MC6 Refer to MC4. The process engineer decides to reduce the size of the test to $\alpha = .01$ and use n=40 chips in the study. What is the probability of a Type II error using an $\alpha = .01$ test when $\sigma = 2.775$?

A. .29

B. .71

C. .90

D. .95

E. .99

MC7 A biologist designs a study to determine if the average wing span of Mexican bats is less than the average wing span of South Carolina bats which is know to be 3.0 mm. She wants to determine how many Mexican bats to include in the study so that the probability of a Type II test at $\mu = 2.8$ mm is at most 0.20 if she uses an α=.05 test. It is well known that the wing spans have a normal distribution. The biologist states she thinks that the wing span of Mexican bats have $\sigma \approx .4$ mm. The sample size must be at least

A. 16

B. 25

C. 35

D. 80

E. cannot be determined with the given information

MC8 You have been assigned to design an experiment to compare the mean responses from a placebo treatment and a new drug. You can either randomly assign n experimental units to each of the treatments or you can pair the $2n$ experimental units based on severity of the disease and then randomly assign the two treatments within each pair of experimental units. Which of the following statements is **TRUE**?

A. Pairing results in a large increase in the power of the paired t-test provided there is negative correlation within the pairs.

B. Pairing results in a large increase in the power of the paired t-test provided there is positive correlation within the pairs.

C. Pairing will always increase the power of the t-test, at least to some degree.

D. Pairing is done to reduce the variance in the difference in the two sample means.

E. Pairing would reduce the power of the t-test if the 2n experimental units have nearly the same level of severity of the disease.