Information about the Final Exam

This is a two hour test!

The Proctor is expected to manually monitor elapsed exam time. If connection problems surface, the Proctor should allow more time as needed. Here is some additional detail.

Open book: YES—Students may access any printed materials, personal computer files and the Canvas course site which includes use of WileyPlus. Use of eBook readers is permitted during the exam provided the reader is resident on the personal computer used during the exam.

Open notes: YES---Any written notes, printed notes or personal computer files.

Open web: NO--- Access to and use of the internet is not permitted with the exception of WileyPlus, which is available via the Canvas course site. No separate portable devices such as Kindles or iPads are permitted during the exam.

Calculators: YES---R, RStudio, Excel or any calculator application which does not require internet access is permitted. Handheld calculators, such as a TI 84, Casio or comparable, are also acceptable.

Scratch Paper: YES

R or RStudio are acceptable computational software. Word processors may be used for essay composition. All word processors are allowed provided they are resident on the student's personal computer. This also includes applications such as Notepad or TextEdit or comparable.

Restroom breaks are not permitted.

Any unresolved questions that arise before or during the exam should be directed to me. Students have a copy of this communication, and have my permission to show it to the proctor.

My Personal Information

Email: Irene.tsapara@northwestern.edu
Personal Email: Irene.tsapara@yahoo.com
Mobile: 312-478-3988

Topics on the Final Exam

- Probability
  - Bayes' Theorem application
  - Calculate probabilities from continuous distributions
  - Calculate means and variances for probability distributions
- Hypothesis Testing
  - Type I and Type II errors
  - Use the T test with a correlation
  - t tests
    - one sample
    - two sample
- Confidence interval construction
- One-way AOV
  - F test
  - p-values
- Calculate linear regression coefficients

The test is two hours, proctored, open book and open notes. The questions are multiple choice. Excel, R or any comparable calculator application may be used. The course site, WileyPlus and files on the computer used for testing are available. No preview of the exam is available. Review questions are available.

The use of portable devices such as kindles and iPads is not allowed unless special arrangements are made. No navigation from the testing site to the internet during the final exam.

Selected Review Problems

Use Bayes' theorem to find the indicated probability.

3) Use the results summarized in the table.

	Approve of mayor	Do not approve of mayor
Republican	8	17
Democrat	18	13
Independent	7	37

One of the 100 test subjects is selected at random. Given that the person selected approves of the mayor, what is the probability that they vote Democrat?

A) 0.674 B) 0.581 C) 0.391 D) 0.545

$$\frac{18/100}{(18/31)(31/100)+(8/25)(25/100)+(7/44)(44/100)} = 18/33 = 0.545$$

Solve the problem.

6) True or False: In a hypothesis test, an increase in α will cause a decrease in the power of the test provided the sample size is kept fixed.

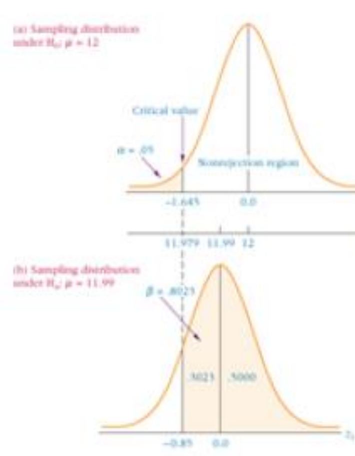
A) True B) False

7) True or False: In a hypothesis test regarding a population mean, the probability of a type II error, β, depends on the true value of the population mean.

A) False B) True

For 6, by increasing the Type I error rate there is a greater chance the null hypothesis will be rejected. This results in an increase in the power.

For 7, power calculations depend on the alternative hypothesis. Refer to Black Section 9.6.



11) The systolic blood pressures of the patients at a hospital are normally distributed with a mean of 138 mm Hg and a standard deviation of 13.5 mm Hg. Find the two blood pressures having these properties: the mean is midway between them and 90% of all blood pressures are between them.

A) 125.9 mm Hg, 150.2 mm Hg
B) 116.8 mm Hg, 161.2 mm Hg
C) 115.8 mm Hg, 160.2 mm Hg
D) 120.7 mm Hg, 155.3 mm Hg

We are looking for an interval that is symmetric with the mean in the middle. To have 90% of the blood pressures between them, 95% of the readings must be to the left of the upper bound, and 5% to the left of the lower bound.

```
> qnorm(0.95, 138, 13.5, lower.tail = TRUE)
[1] 160.2055
> qnorm(0.05, 138, 13.5, lower.tail = TRUE)
[1] 115.7945
```

Construct the indicated confidence interval for the difference between population proportions $p_1 - p_2$. Assume that the samples are independent and that they have been randomly selected.

13) $x_1 = 15$, $n_1 = 50$ and $x_2 = 23$, $n_2 = 60$; Construct a 90% confidence interval for the difference between population proportions $p_1 - p_2$.

A) $-0.232 < p_1 - p_2 < 0.065$
B) $0.477 < p_1 - p_2 < 0.122$
C) $0.123 < p_1 - p_2 < 0.477$
D) $0.151 < p_1 - p_2 < 0.449$

This can be done conveniently in R.

```
> x <- c(15, 35)
> y <- c(23, 37)
> M <- rbind(x,y)
> prop.test(M, alternative = c("two.sided"), conf.level = 0.90, correct = FAL
SE)

2-sample test for equality of proportions without continuity correction

data:  M
X-squared = 0.8376, df = 1, p-value = 0.3601
alternative hypothesis: two.sided
90 percent confidence interval:
 -0.23173354  0.06506688
sample estimates:
   prop 1    prop 2
0.3000000 0.3833333
```

Perform the indicated hypothesis test. Assume that the two samples are independent simple random samples selected from normally distributed populations. Also assume that the population standard deviations are equal ($\sigma_1 = \sigma_2$), so that the standard error of the difference between means is obtained by pooling the sample variances.

14) A researcher was interested in comparing the amount of time spent watching television by women and by men. Independent simple random samples of 14 women and 17 men were selected, and each person was asked how many hours he or she had watched television during the previous week. The summary statistics are as follows.

Women	Men
$\bar{x}_1 = 11.4$ hr	$\bar{x}_2 = 16.8$ hr
$s_1 = 4.1$ hr	$s_2 = 4.7$ hr
$n_1 = 14$	$n_2 = 17$

Use a 0.05 significance level to test the claim that the mean amount of time spent watching television by women is smaller than the mean amount of time spent watching television by men. Use the traditional method of hypothesis testing.

For this problem, a one-sided test is required. The alternative will be framed as a positive.

```
> s1.2 <- 4.1^2
> s2.2 <- 4.7^2
> n1 <- 14
> n2 <- 17
> pool <- sqrt((s1.2*(n1-1)+s2.2*(n2-1))/(n1+n2-2))
> den <- pool*sqrt(1/n1+1/n2)
> x1 <- 11.4
> x2 <- 16.8
> t <- (x2-x1)/den
> t
[1] 3.369099
> pt(t,29,lower.tail=FALSE)
[1] 0.001073162
> qt(0.95,29,lower.tail=TRUE)
[1] 1.699127
```

Use the given data to find the equation of the regression line. Round the final values to three significant digits, if necessary.

15)

x	6	8	20	28	36
y	2	4	13	20	30

A) $\hat{y} = -2.79 + 0.897x$

B) $\hat{y} = -2.79 + 0.950x$

C) $\hat{y} = -3.79 + 0.801x$

D) $\hat{y} = -3.79 + 0.897x$

```
> x <- c(6, 8, 20, 28, 36)
> y <- c(2, 4, 13, 20, 30)
> lm(y~x)

Call:  lm(formula = y ~ x)

Coefficients:
(Intercept)            x
    -3.7900       0.8975
```

16) For the data below, determine the value of the linear correlation coefficient r between y and x^2.

x	1.2	2.7	4.4	6.6	9.5
y	1.6	4.7	9.9	24.5	39.0

A) 0.913 B) 0.990 C) 0.873 D) 0.985

```
> x <- c(1.2, 2.7, 4.4, 6.6, 9.5)
> y <- c(1.6, 4.7, 9.9, 24.5, 39.0)
> x <- x^2
> cor(x, y, method = c("pearson"))
[1] 0.9902759
```

---

18) In studying the occurrence of genetic characteristics, the following sample data were obtained. At the 0.05 significance level, test the claim that the characteristics occur with the same frequency.

Characteristic	A	B	C	D	E	F
Frequency	28	30	45	48	38	39

This is a Chi-square goodness-of-fit test. The counts are expected to be equal under the null hypothesis. This expectation is 38 which needed to be compared against the observed counts.

```
> obs <- c(28, 30, 45, 48, 38, 39)
> ec <- rep(sum(obs)/6, times = 6)
> diff <- sum((obs - ec)^2/ec)
> diff
[1] 8.263158
> pchisq(diff, df = 5, lower.tail = FALSE)
[1] 0.1423164
> qchisq(0.95, df = 5, lower.tail = TRUE)
[1] 11.0705
```

---

20) Fill in the missing entries in the following partially completed one-way ANOVA table.

Source	df	SS	MS=SS/df	F-statistic
Treatment	3			11.16
Error		13.72	0.686	
Total				

Error degrees of freedom = 13.72/0.686 = 20. Total degrees of freedom = 3 + 20 = 23

Treatment MS = 11.16(.686) = 7.656. Treatment SS = 7.656(3) = 22.97

Total SS = 13.72 + 22.97 = 36.69

Problem from Black Tables 11.2 and 11.3

TABLE 11.2 **Valve Openings by Operator**

1	2	3	4
6.33	6.26	6.44	6.29
6.26	6.36	6.38	6.23
6.31	6.23	6.58	6.19
6.29	6.27	6.54	6.21
6.40	6.19	6.56	
	6.50	6.34	
	6.19	6.58	
	6.22		

```
> y <- c(6.33,6.26,6.31,6.29,6.4,6.26,6.36,6.23,6.27,6.19,6.5,6.19,6.22,6.44,6.38,
+        6.58,6.54,6.56,6.34,6.58,6.29,6.23,6.19,6.21)
> x <- c(rep(1,times=5),rep(2,times=8),rep(3,times=7),rep(4,times=4))
> valve <- data.frame(cbind(y,x))
> valve$x <- factor(valve$x)
> str(valve)
'data.frame':   24 obs. of  2 variables:
 $ y: num  6.33 6.26 6.31 6.29 6.4 6.26 6.36 6.23 6.27 6.19 ...
 $ x: Factor w/ 4 levels "1","2","3","4": 1 1 1 1 1 2 2 2 2 2 ...
> results <- aov(y~x,data=valve)
> summary(results)
            Df Sum Sq Mean Sq F value   Pr(>F)
x            3 0.2366 0.07886   10.18 0.000279 ***
Residuals   20 0.1549 0.00775

> aggregate(y ~ x, data = valve, mean)
  x        y
1 1 6.318000
2 2 6.277500
3 3 6.488571
4 4 6.230000
> TukeyHSD(results, conf.level = 0.95)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = y ~ x, data = valve)

$x
          diff         lwr         upr     p adj
2-1 -0.0405000 -0.18093243  0.09993243 0.8502854
3-1  0.1705714  0.02633255  0.31481031 0.0169205
4-1 -0.0880000 -0.25324639  0.07724639 0.4613461
3-2  0.2110714  0.08358107  0.33856179 0.0008519
4-2 -0.0475000 -0.19834863  0.10334863 0.8144408
4-3 -0.2585714 -0.41296992 -0.10417294 0.0007541
```

z-test on proportions from two populations

Use the traditional method to test the given hypothesis. Assume the samples are independent and that they have been randomly selected. Use the given sample data to test the claim that $p_1 < p_2$. Use a significance level of 0.10.

Sample 1	Sample 2
$n_1 = 462$	$n_2 = 380$
$x_1 = 84$	$x_2 = 95$

$$H_0: p_1 - p_2 = 0$$
$$H_A: p_2 - p_1 > 0$$

$$z = \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)}{\sqrt{(\tilde{p}(1-\tilde{p})\left(\frac{1}{n_2} + \frac{1}{n_1}\right))}}$$

Reject the null hypothesis. 2.406324 > 1.281552

p-value = 0.00805698 < 0.1

Using R and prop.test()

```
> s1 <- c(84, 462)
> s2 <- c(95, 380)
> s12 <- rbind(s1,s2)
> s12[,2] <- s12[,2] - s12[,1]
> s12
   [,1] [,2]
s1   84  378
s2   95  285
> prop.test(s12,alternative = "less", conf.level = 0.95, correct = FALSE)

2-sample test for equality of proportions without continuity correction

data:  s12
X-squared = 5.7904, df = 1, p-value = 0.008057
alternative hypothesis: less
```

Note: sqrt(5.7904) = 2.406325