NAME: ANSWER Key

Directions: Be sure to show all of your work. An answer alone will not receive any credit. You must show a formula or how you arrived at your answer. Partial credit will be given on all problems.

10 1. In an article in the Cincinnati *Enquirier* earlier this year, data was presented on the total number of school bus crashes in each of 88 counties in Ohio in the 2005-06 school year. They also report the number of accidents per millions of bus miles driven. Below is the Minitab descriptive statistics output for this last variable. Use this information to answer the questions below.

```
Variable            N   Mean   SE Mean  StDev     Minimum      Q1   Median     Q3  Maximum
Accidents per mi   88  6.696    0.437  4.097  0.000000000  3.543   6.190  8.413   22.340
```

-3 if Z-scores 3.818 3σ

5 a. Suppose I ask you to obtain the Modified Boxplot. Without actually drawing the boxplot, are there any outliers in this data set?

10/10

upper fence $= Q_3 + 1.5(IQR) = 8.413 + 1.5(\underbrace{8.413-3.543}_{4.87}) = 8.413 + 7.305 = 15.718$ ($Q_3 - Q_1$)

lower fence $= Q_1 - 1.5\, IQR = 3.543 - 7.305 = -3.7620 = 0$

No values less than 0 so no outliers in low direction.

but max $= 22.340 > 15.718$ So at least one outlier on high side.

5 b. In which direction is the data skewed? Give **two** pieces of evidence from the information above that supports your claim.

skewed POSITIVELY since (1) mean > median (6.696, 6.190)

(2) $Q_1 - \min = 3.543 - 0 = 3.543 < Q_3 - \max = 13.927$

10 2. The following data represent two RS's of the running times (in mins) of films produced by 2 motion-picture companies:

Company	Movie 1	2	3	4	5	6	7	n	Avg	SD	90% CI	Two Sample 90% CI
A	102	86	98	109	92			5	97.4	8.88	(88.94, 105.86)	(-35.51, 10.31)
B	81	165	97	134	92	87	114	7	110	30.22	(87.80, 132.20)	
Difference A - B	21	-79	1	-25	0	-87	-114	5	-40.43	52.33	(-78.86, -1.99)	

-5

Using a 90% confidence interval determine if Company A's movies run 10 mins longer than Company B, on the average. Explain what you used and how you used it!

-3

10/10

Since samples are independent (Two different RS's). use Two Sample CI (-35.51, 10.31). So with 90% conf we can conclude that the difference in mean run time btw A and B is btw -35.51 + 10.31. So company A's movie run up to 35 shorter than B to possibly as much as 10.31 min longer than B's. So it is possible that A's movie run 10 min longer than B

15 3. In a study of the effect of population size in various cities in the US on ozone concentrations, a random sample of 10 cities was selected and the ozone measured (in parts per billion) as was the population size (in millions). Using the attached SAS regression output, answer the questions that follow.

4 a. What is the estimated regression relationship between Ozone concentrations and population size?

Predicted Ozone = 125.93386 + 1.65412 · Popln Size

1

7 b. Obtain a 95% confidence interval for β_1 and interpret your answer using the words of this problem.

15/15

$b_1 \pm t_{(.025, 8)}\, se(b_1)$ 2 1.65 ± 2.306 (0.402) 1.65 ± 0.9270 [.723, 2.577] 1 1 1 1

with 95% conf we can conclude that for each million increase in popln size Ozone conc goes up on average btw 0.72 to 2.58 ppb. 2

4 c. What is the estimated correlation between ozone and population size for US cities?

$r = \text{sign of } b_1 \sqrt{R^2} = +\sqrt{0.6791} = 0.8241$

R^2 -2

15 4. It is suspected that the environmental temperature in which industrial batteries are activated affects their life. Thirty homogeneous batteries were tested, six at each of five temperatures, and the lifetime (in days) of the batteries measured. Use the SAS output attached to answer the questions below.

5 a. Can we conclude, with 95% confidence, that the batteries are different? **What** can you conclude **AND why** can you conclude what you do?

Anova f test is signif $f^* = 69.05$ $p < 0.0001$ so we can conclude with 95% conf that the mean lifetimes of batteries is different for the different activation temps. 1

15/15

5 b. Can you conclude that there is an optimum temperature at which batteries should be activated to maximize lifetimes? Explain.

since CI's for $\mu_{50} - \mu_{75}$ includes zero μ_{50} is not diff from μ_{75} but μ_{50} & μ_{75} CI's versus other temps are all positive hence max lifetime is for a Temp of 50 or 75 2

5 c. Of the assumptions in ANOVA, which is most seriously in question in this analysis? Why?

1 Normality – stem/leaf ≈ Normal + Normal Prob Plot ≈ straight so Normal ok.

4 Constant Var – questionable since Temp 25 has a very small spread compared to 50 so serious questions on Const. var

[5] 5. A random sample of 120 is taken from a population with unknown mean but whose variance is approximately 5. What is the probability our sample average will be within 0.5 of the population mean?

$n = 120$ from popln $(\mu, \sigma^2 \doteq 5)$ n large $\Rightarrow \bar{X} \approx N(\mu, \sigma^2/n = 5/120)$

$\frac{5}{5}$

$$P(\mu - 0.5 \le \bar{X} \le \mu + 0.5) \doteq P\left(\frac{\mu - 0.5 - \mu}{\sqrt{5/120}} \le Z \le \frac{(\mu + 0.5) - \mu}{\sqrt{5/120}} \right)$$

$$\doteq P(-2.45 \le Z \le 2.45) = P(Z < 2.45) - P(Z < -2.45)$$

$$= 0.9929 - 0.0071 = 0.9858$$

[10] 6. An experiment consists of tossing three fair dice. Let a Success occur if a die comes up 1 or 2 and consider a 3, 4, 5, or 6 a Failure. Let X be the number of Successes MINUS the number of Failures on the three dice. What is the probability function of X?

$P(S) = \frac{2}{6} = \frac{1}{3}$

$P(F) = \frac{4}{6} = \frac{2}{3}$

$\frac{10}{10}$

DIE 1	2	3	(Prob)	X #S−#F
S	S	S	$\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{1}{3} = 1/27$	$3-0 = 3$
S	S	F	$1/3 \cdot 1/3 \cdot 2/3 = 2/27$	$2-1 = 1$
S	F	S	$1/3 \cdot 2/3 \cdot 1/3 = 2/27$	$2-1 = 1$
S	F	F	$1/3 \cdot 2/3 \cdot 2/3 = 4/27$	$1-2 = -1$
F	S	S	$2/3 \cdot 1/3 \cdot 1/3 = 2/27$	$2-1 = 1$
F	S	F	$2/3 \cdot 1/3 \cdot 2/3 = 4/27$	$1-2 = -1$
F	F	S	$2/3 \cdot 2/3 \cdot 1/3 = 4/27$	$1-2 = -1$
F	F	F	$2/3 \cdot 2/3 \cdot 2/3 = 8/27$	$0-3 = -3$

−2 if 1 is just SSF

x	$f(x)$	
3	1/27	.0370
1	6/27	.2222
−1	12/27	.4444
−3	8/27	.2963
	27/27	

[10] 7. An important system acts in support of a vehicle in our space program. A single crucial component works only 85% of the time. In order to enhance the reliability of the system, it is decided that 3 components will be installed in parallel such that the system fails only if they all fail. Assume the components act independently and that they are equivalent in the sense that all 3 of them have an 85% success rate. ~~Consider the random variable X = number of components out of 3 that fail~~. If the desire is to have the system be successful with probability 0.99, are the three components sufficient? If not, how many are required?

$Sys = C_1 \cap C_2 \cap C_3$ $P(\text{System Fail}) = P(C_1 \text{ fail} \cap C_2 \text{ fail} \cap C_3 \text{ fail})$

$P(C_i \text{ fail}) = 1 - .85 = .15$

$\frac{10}{10}$

$P(\text{System Works}) = 0.99$

"$1 - P_r(\text{system Fails})$" $= 1 - P(C_1 \text{ fails} \cap C_2 \text{ fails} \cap C_3 \text{ fails})$

$= 1 - P(C_1 \text{ fails}) \cdot P(C_2 \text{ fail}) \cdot P(C_3 \text{ fail})$ b/c indep

$= 1 - 0.15(0.15)(0.15)$

$= 1 - 0.0034 = .9966 > 0.99$ so 3 are sufficient

[10] 8. Using the information in the following article, determine whether the proportions of ties contaminated with strains of bacteria are different for medical staff and security guards. Use $\alpha = 0.05$. Give the 0 to 7 steps of a hypothesis test as we did in class!

> **Is Your Doctor's Necktie Contaminated?** By Nissa Simon *AARP Bulletin*, July-August 2004
> *If your doctor's tie is making you sick, the problem might be more than his affinity for paisley. Doctors are likely to acquire disease-causing bacteria on their neckties as they move from patient to patient. Researchers at New York Hospital Medical Center of Queens swabbed 42 ties worn by medical personnel and 10 ties worn by security guards who had little contact with patients. They grew microorganisms from each tie. Twenty five of the ties of the medical staff carried several strains of bacteria that could cause infections; the tie of only one security guard was contaminated. Although there's no real danger, says Stuart Levy, president of the Alliance for the Prudent Use of Antibiotics, "maybe we should all wear bow ties. I do."*

1 0. P_M = Prop of ties of med staff with strains of bacteria a diff

1 P_S = " " " sec. " " "

1 1. H_0: $P_M = P_S$ $\equiv$ Staff designation + whether tie has bacteria are indep

1 2. H_A: $P_M \neq P_S$ = staff + bac are dep.

3. $\alpha = 0.05$

10/10

4. Exp: $\frac{26 \cdot 42}{52} = 21 \geq 5$

$\frac{26 \cdot 42}{52} = 21 \geq 5$

$\frac{26 \cdot 10}{52} = 5 \geq 5$

$\frac{26 \cdot 10}{52} = 5 \geq 5$

3

so n large enough

Expected Count

	Bacteria	No Bact	
Med	25/21	17/21	42
Sec	1/5	9/5	10
	26	26	52

$$\chi^2 = \frac{(25-21)^2}{21} + \frac{(17-21)^2}{21} + \frac{(1-5)^2}{5} + \frac{(9-5)^2}{5}$$

1

$$= \frac{16}{21} + \frac{16}{21} + \frac{16}{5} + \frac{16}{5} = \frac{32}{21} + \frac{32}{5} = 7.9238$$

1 5. p-value = $P(\chi^2_{(1)} > 7.9238) < 0.005$

1 6. Reject H_0 since $p = 0.005 < .05$

1 7. With 95% conf we can conclude that the prop of ties with bact strains is different for med staff + sec staff.

7 9. The probability distribution of X, the number of flaws per 10 meters of a synthetic fabric in continuous rolls of uniform width, is given by:

$C = 15x + 100$

x	0	1	2	3	4
$f_X(x)$	0.41	0.37	0.16	0.05	0.01
c	100	115	130	145	160
c.f	41	42.55	20.8	7.25	1.6
$c^2.f$	4100	4893.25	2704	1051.25	256

$E(C) = \Sigma = 113.2$ 2
$E(C^2) = 13004.5$ 2

If the mean of X is 0.88 and if the cost to the manufacturer to correct these flaws depends on the number of flaws per 10 meters as follows, Cost = C = \$15 * X + \$100, what is the **variance of the cost to correct flaws in this fabric**?

7/7

$V(c) = V(15X+100) = 15^2 V(x) = 15^2\left(E(x^2) - \mu_x^2\right)$ 2 2 $\mu_x = .88$

$E(x^2) = 0^2(.41) + 1^2(.37) + 2^2(.16) + 3^2(.05) + 4^2(.01)$

2 $= 0 + .37 + 0.64 + 0.45 + .16 = 1.62$

1 so $V(c) = 15^2(1.62 - .88^2) = 225(0.8456) = 190.26$

OR $V(C) = E(C^2) - E(C)^2 = 13004.5 - 113.2^2 = 190.26$ ✓
2 $113.2^2 = 12814.24$ 1

8 10. The random variable X has pdf given by $f_X(x) = \frac{200}{3}\frac{1}{(x+5)^3}$, for $0 < x < 5$. Find the mean of X.

Hint: Find the expected value of X + ~~6~~ 5

$E(X+5) = \int_0^5 (x+5)\left[\frac{200}{3}\frac{1}{(x+5)^3}\,dx\right] = \frac{200}{3}\int_0^5 (x+5)^{-2}\,dx$ 2

8/8

$= \frac{200}{3}\frac{(x+5)^{-1}}{-1}\Big|_0^5 = \frac{-200}{3}\left[(5+5)^{-1} - (0+5)^{-1}\right] = -\frac{200}{3}\left[\frac{1}{10} - \frac{1}{5}\right]$ 2

$= -\frac{200}{3}\left[\frac{1}{10} - \frac{2}{10}\right] = -\frac{200}{3}\left[-\frac{1}{10}\right] = \frac{20}{3} = E(X+5)$ so $= E(X) + 5$ 2

$\therefore E(X) = \frac{20}{3} - 5 = \frac{5}{3}$ ✓ 2

SAS OUTPUT FOR PROBLEM #3

```
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
Final Exam Winter 2007.sas                                                                  1
Exercise 11.46 and 13.23 in WMMY 8th

Obs    OZONE    POPLN

  1     126      0.6
  2     135      4.9
  3     124      0.2
  4     128      0.5
  5     130      1.1
  6     128      0.1
  7     126      1.1
  8     128      2.3
  9     128      0.6
 10     129      2.3
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
Final Exam Winter 2007.sas                                                                  2
Exercise 11.46 and 13.23 in WMMY 8th

The REG Procedure
Model: MODEL1
Dependent Variable: OZONE

Number of Observations Read          10
Number of Observations Used          10


                         Analysis of Variance

                                   Sum of           Mean
Source                   DF       Squares         Square    F Value    Pr > F

Model                     1      52.70025       52.70025      16.93    0.0034
Error                     8      24.89975        3.11247
Corrected Total           9      77.60000


Root MSE              1.76422    R-Square     0.6791
Dependent Mean      128.20000    Adj R-Sq     0.6390
Coeff Var             1.37615


                      Parameter Estimates

                      Parameter       Standard
Variable      DF       Estimate          Error    t Value    Pr > |t|

Intercept      1      125.93386        0.78393     160.64      <.0001
POPLN          1        1.65412        0.40199       4.11      0.0034
```

SAS OUTPUT FOR PROBLEM #4

++

```
Final Exam Winter 2007.sas                                                                                      3
Exercise 11.46 and 13.23 in WMMY 8th
```

TEMP (°C)	LIFETIME					
0	55	55	57	54	54	56
25	60	61	60	60	60	60
50	70	72	72	68	77	77
75	72	72	72	70	68	69
100	65	66	60	64	65	65

++

```
Final Exam Winter 2007.sas                                                                                      4
Exercise 11.46 and 13.23 in WMMY 8th

The GLM Procedure

        Class Level Information

Class          Levels    Values

TEMP                5    0 25 50 75 100


Number of Observations Read          30
Number of Observations Used          30
```

++

```
Final Exam Winter 2007.sas                                                                                      5
Exercise 11.46 and 13.23 in WMMY 8th

The GLM Procedure

Dependent Variable: LIFETIME

                                          Sum of
Source                       DF          Squares     Mean Square    F Value    Pr > F

Model                         4      1252.133333      313.033333      69.05    <.0001

Error                        25       113.333333        4.533333

Corrected Total              29      1365.466667


R-Square     Coeff Var      Root MSE    LIFETIME Mean

0.917000      3.299322      2.129163         64.53333


Source                       DF        Type I SS     Mean Square    F Value    Pr > F

TEMP                          4      1252.133333      313.033333      69.05    <.0001


Source                       DF      Type III SS     Mean Square    F Value    Pr > F

TEMP                          4      1252.133333      313.033333      69.05    <.0001
```

The GLM Procedure

Bonferroni (Dunn) t Tests for LIFETIME

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

Alpha	0.01
Error Degrees of Freedom	25
Error Mean Square	4.533333
Critical Value of t	3.72514
Minimum Significant Difference	4.5792

Means with the same letter are not significantly different.

Bon Grouping	Mean	N	TEMP
A	72.667	6	50
A			
A	70.500	6	75
B	64.167	6	100
B			
B	60.167	6	25
C	55.167	6	0

++

The GLM Procedure

Bonferroni (Dunn) t Tests for LIFETIME

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than Tukey's for all pairwise comparisons.

Alpha	0.01
Error Degrees of Freedom	25
Error Mean Square	4.533333
Critical Value of t	3.72514
Minimum Significant Difference	4.5792

Comparisons significant at the 0.01 level are indicated by ***.

TEMP Comparison	Difference Between Means	Simultaneous 99% Confidence Limits		
50 - 75	2.167	-2.413	6.746	
50 - 100	8.500	3.921	13.079	***
50 - 25	12.500	7.921	17.079	***
50 - 0	17.500	12.921	22.079	***
75 - 50	-2.167	-6.746	2.413	
75 - 100	6.333	1.754	10.913	***
75 - 25	10.333	5.754	14.913	***
75 - 0	15.333	10.754	19.913	***
100 - 50	-8.500	-13.079	-3.921	***
100 - 75	-6.333	-10.913	-1.754	***
100 - 25	4.000	-0.579	8.579	
100 - 0	9.000	4.421	13.579	***
25 - 50	-12.500	-17.079	-7.921	***
25 - 75	-10.333	-14.913	-5.754	***
25 - 100	-4.000	-8.579	0.579	
25 - 0	5.000	0.421	9.579	***
0 - 50	-17.500	-22.079	-12.921	***
0 - 75	-15.333	-19.913	-10.754	***
0 - 100	-9.000	-13.579	-4.421	***
0 - 25	-5.000	-9.579	-0.421	***

```
Final Exam Winter 2007.sas                                                                    8
Exercise 11.46 and 13.23 in WMMY 8th

The UNIVARIATE Procedure
Variable:  R

   Stem Leaf                     #  Boxplot                       Normal Probability Plot
      4 33                       2     0        4.5+                                           *   *+++++
      3                                            |                                          +++++
      2                                            |                                     +++++
      1 55588                    5     |        1.5+                                ++*** *
      0 88888                    5  +--+--+        |                           +++*****
     -0 77522222222222          11  *-----*        |                     ****+*+* *
     -1 522                      3     |       -1.5+                 **+++
     -2 75                       2     |           |            +*+*+
     -3                                            |       +++++
     -4 72                       2     0       -4.5++++++*   *
        ----+----+----+----+                       +----+----+----+----+----+----+----+----+----+----+
                                                        -2        -1         0        +1        +2
```