# Stat 252 – Lab Exam 1 Solutions (Practice)

**Case Study 1 (15 marks): Use 252labexam1a to answer the following questions.**

In an industrial laboratory, under uniform conditions, batches of electrical insulating fluid were subjected to constant voltages until the insulating property of the fluids broke down. Five different voltage levels (24, 26, 28, 30, and 32 kV) were studied. The measured responses were the times, in minutes, until breakdown. A natural log transformation was taken for breakdown time.

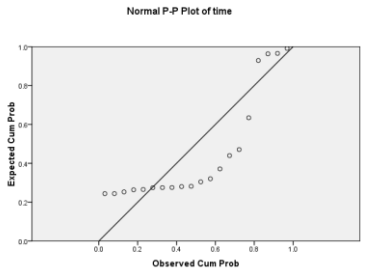
The data file contains the following variables:

*voltage*: in kV
*group*: a grouping code for voltage (1 = 24 kV, 2 = 26 kV, 3 = 28 kV, 4 = 30 kV, 5 = 32 kV)
*time:* breakdown time
*logtime*: the natural log of breakdown time

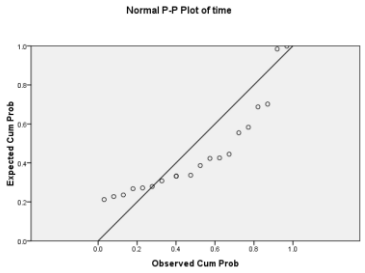
1. The sample standard deviations of breakdown time for the five groups are, respectively:
   a) 259.92, 244.30, 61.77, 45.17, 7.87
   **b) 345.53, 363.39, 41.15, 60.04, 9.08**
   c) 77.26, 81.26, 9.20, 13.43, 2.03
   d) 18.59, 19.06, 6.41, 7.75, 3.01
   e) 1.29, 1.04, 0.73, 1.29, 1.02, 1.64

2. One of the plots below is the normal probability plot of standardized residuals of breakdown time at 32 kV. Which plot is it?

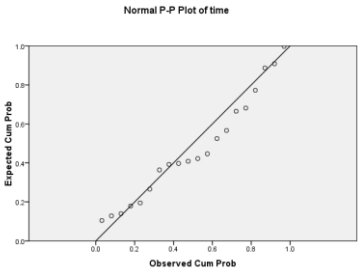
a)



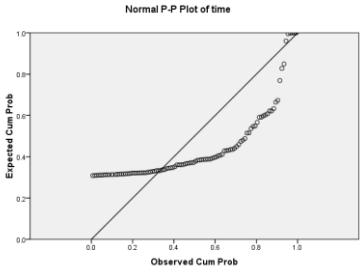

**b)**



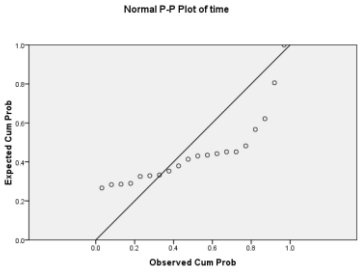

c)




d)



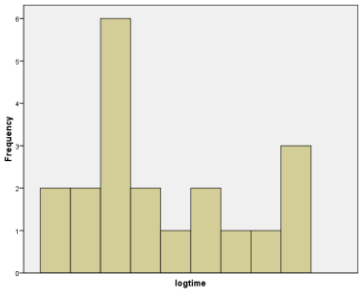

e)



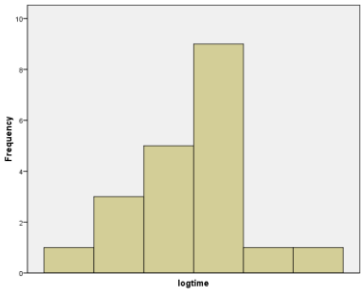

3. Consider a One-Factor Model with breakdown time as the response. Which of the following best describes the assumptions of normality and equal variability?
   a) Neither assumption appears to be violated.
   b) Only equal variability is violated.
   c) Normality is perfectly satisfied but equal variability is violated.
   d) Only normality is violated.
   **e) Both assumptions are violated.**

4. One of the plots below is the histogram of the natural log of breakdown time at 28 kV. Which plot is it?

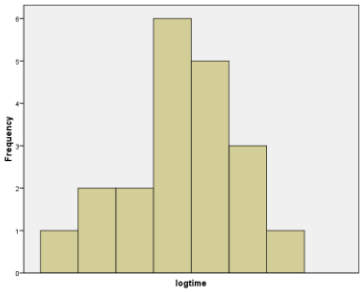
a)



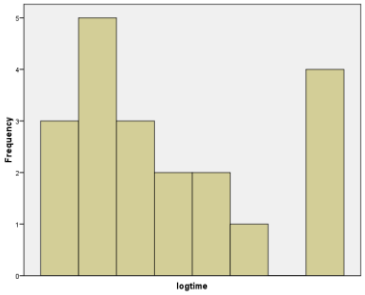

b)



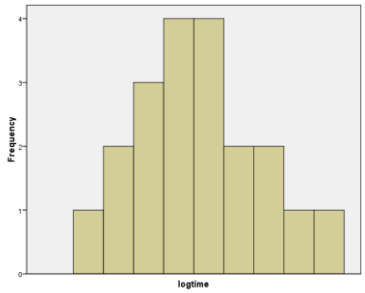

**c)**




d)




e)




5. Consider a One-Factor Model with the natural log of breakdown time as the response. What is the best estimate for the common standard deviation of the five groups?
a) 1.64 b) 1.44 c) 0.16 d) 1.20 **e) 1.10**

6. Consider a One-Factor Model with the natural log of breakdown time as the response. In the test for any mean differences, what is the distribution of the test statistic under the null hypothesis?
a) F(94, 5) b) F(95, 99) c) F(4, 99) **d) F(4, 95)** e) F(5, 100)

7. Consider a One-Factor Model with the natural log of breakdown time as the response. In the test for any mean differences, what is the p-value?
**a) Less than 0.001.**
b) Between 0.001 and 0.01.
c) Between 0.01 and 0.05.
d) Between 0.05 and 0.1.
e) Greater than 0.1.

8. Consider a One-Factor Model with the natural log of breakdown time as the response. In the test for any mean differences, what are the sum-of-squared residuals for the models under the null hypothesis and the alternative hypothesis, respectively?
**a) 267.533 and 114.120**
b) 153.433 and 114.120
c) 114.120 and 153.433
d) 153.433 and 267.533
e) 267.553 and 153.433

9. Consider the following contrast: the average mean logged breakdown time at 24, 26, and 28 kV vs. the average mean logged breakdown time at 30 and 32 kV. What is the estimate of this contrast?
**a) 2.24** b) 3.01 c) 6.73 d) 9.02 e) 13.46

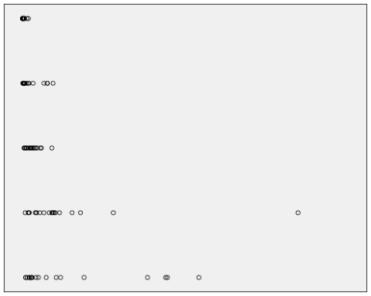
10. What is the standard error for the difference in mean logged breakdown time at 26 kV vs. 24 kV?
a) 0.174 b) 0.371 **c) 0.347** d) 0.694 e) 0.742

11. What is the p-value for the test to determine if mean logged breakdown time is higher at 26 kV vs. 24 kV?
a) 1.144 b) 0.599 c) 0.572 d) 0.300 **e) 0.286**

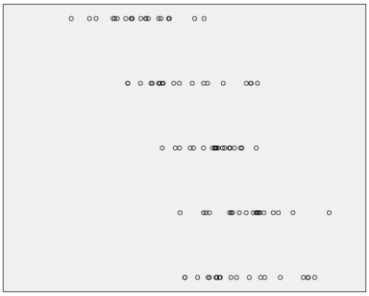
Now, consider voltage to be a numerical explanatory variable and fit a simple linear regression of mean logged breakdown time on voltage.

12. Which scatterplot below is the scatterplot of logged breakdown time vs. voltage?

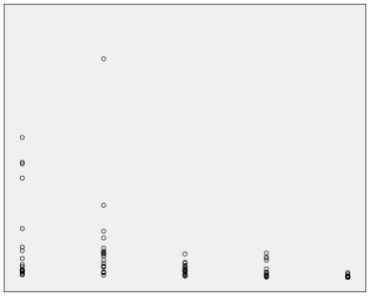
a)



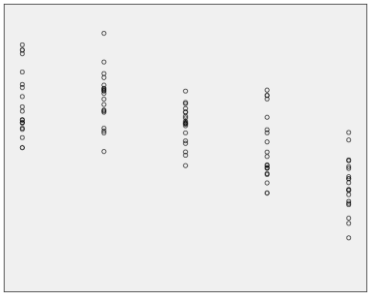
b)



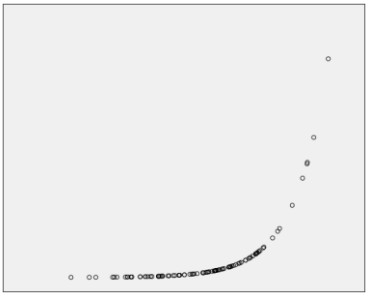
c)



**d)**



e)



13. What are the sum-of-squared residuals and the degrees of freedom for the simple linear regression model?
    a) 7.467 and 7
    b) 114.12 and 95
    c) 137.51 and 1
    **d) 130.04 and 98**
    e) 267.55 and 99

14. What is the estimated multiplicative effect of an additional 2 volts on the median breakdown time?
    a) 2.293
    b) 0.660
    **c) 0.436**
    d) -0.415
    e) -0.830

15. For the t-test to determine if there is a significant linear relationship between voltage and mean logged breakdown time, what are the degrees of freedom for the test statistic under the null hypothesis?
    a) 1 b) 7 c) 95 **d) 98** e) 99

**Case Study 2 (15 marks): Use 252labexam1b to answer the following questions.**

You are the Dean of a highly respected mathematics department at a large university, accepting thousands of applications, and want to admit only those students who will successfully complete their first year with a GPA of at least 3.0. In order to predict a student's GPA at the end of their first year, you decide to consider the mathematics portion of the student's SAT score, the reading comprehension portion of the student's SAT score, and gender as predictors (or explanatory) variables.

The data file contains the following variables:
*gpa*: the student's GPA at the end of their first year university
*satmath*: the student's SAT math score in hundreds (example: satmath = 3.00 implies the student scored 300)
*satread*: the student's SAT reading comprehension score in hundreds
*female*: female = 1 for a female student and female = 0 for a male student
*satint = satmath*satread*: The product (interaction) of *satmath* and *satread*

First, you want to determine if there is any significant difference in mean GPA for males and females. For questions 16 and 17, only consider gender as a factor.

16. What is the p-value for the test to determine if there is any difference in mean GPA for males and females?
    a) 0.000 b) 0.033 **c) 0.066** d) 0.068 e) 0.132

17. What is the 95% confidence interval for the mean difference in mean GPA for females vs. males?
    a) (-0.940, 0.032)
    **b) (-0.032, 0.940)**
    c) (2.081, 2.769)
    d) (-2.769, -2.081)
    e) (-0.947, 0.034)

You believe there is a positive, straight-line relationship between mean gpa and satmath. You consider the following model (Model 1). Use it to answer questions 18 – 22. To make sure you have the correct output, the F-statistic for the test of linear significance is 22.693.

Model 1: $\mu(gpa \mid satmath) = \beta_0 + \beta_1 satmath$

18. Fit the simple linear regression model of mean gpa vs. satmath. The estimated regression equation is:
    a) 1.552 + 0.246*satmath*
    **b) 1.552 + 0.255*satmath***
    c) 0.246 + 0.054*satmath*
    d) 0.246 + 0.255*satmath*
    e) 0.255 + 0.054*satmath*

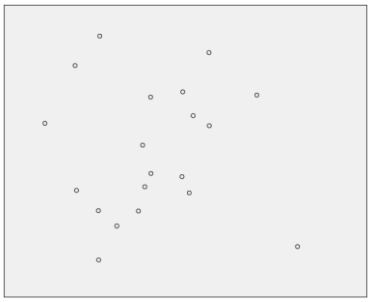
19. The estimated correlation between gpa and satmath is:
    a) 0.379 b) 0.533 c) 0.255 d) 0.558 **e) 0.747**

20. The 95% confidence interval for the mean gpa that is associated with a score of 500 on the math portion of the SAT is:
    a) (1.035, 2.069)
    b) (0.143, 0.368)
    c) (1.750, 3.909)
    **d) (2.633, 3.022)**
    e) (2.007, 3.648)

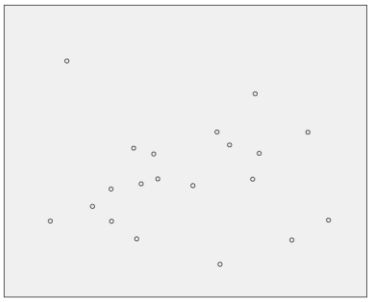
21. What is the estimated effect on mean gpa of an additional 50 points on satmath?
    a) 14.302 b) 0.255 c) 12.75 **d) 0.1275** e) 1.6795

22. Which plot below is the scatterplot of the standardized residuals vs. the standardized predicted values?

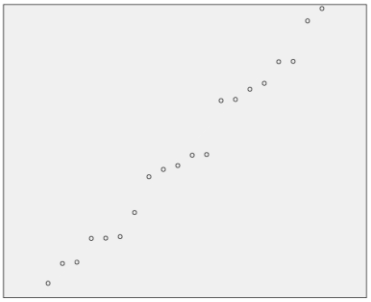
a)



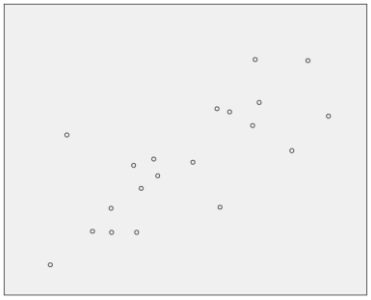
**b)**



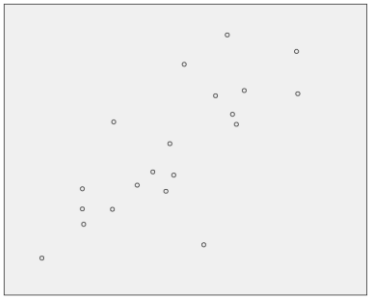
c)



d)



e)



To try and improve your model, you decide to include the student's SAT reading comprehension score and their gender. As well, you will also consider an interaction term between satmath and satread. "You've" defined the model below (Model 2). Use it to answer questions 23 – 27. To make sure you have the correct output, the standard error for the model below is 0.3125.

Model 2:
$$\mu(gpa \mid satmath, satread, GENDER) = \beta_0 + \beta_1 satmath + \beta_2 satread + \beta_3\, female + \beta_4 (satmath \times satread)$$

23. The percentage in variation accounted for by the regression model above is:
a) 0.866 **b) 0.750** c) 0.683 d) 0.313 e) 0.563

24. What would you estimate is the mean gpa for a female student who scored 500 on both portions of the SAT to be?
a) 2.749 b) 3.224 c) 3.370 **d) 2.895** e) 2.755

25. Does the effect of satmath depend on satread? What is the p-value for the test to answer this question?
a) 0.000 b) 0.139 c) 0.175 d) 0.229 **e) 0.597**

26. You wish to carry out a test to determine if mean gpa is higher for females than for males, after accounting for a satmath, satread, and their interaction. The p-value for this test is:
a) 0.000 b) 0.033 c) 0.066 **d) 0.242** e) 0.483

27. Which statement best describes the model?
a) No factors appear to be significant.
**b) Some factors appear to be significant.**
c) All factors appear to be significant.
d) The p-values reported are unreliable as there is a violation in the normality assumption.
e) The p-values reported are unreliable as there is a violation in the constant variability assumption.

Now you wish to compare model 1 to model 2 to determine which is best.

28. In order to compare model 1 and model 2, you have to identify the sum-of-squared residuals and degrees of freedom for the reduced model and the full model. Let SSR(r) and SSR(f) denote the sum-of-squared residuals for the reduced and full models, respectively. Let df(r) and df(f) denote the degrees of freedom for the reduced and full models, respectively. The four values (SSR(r), df(r), SSR(f), df(f)) are:
    a) 3.266, 1, 4.391, 4
    **b) 2.590, 18, 1.465, 15**
    c) 5.856, 19, 4.391, 4
    d) 5.856, 19, 1.465, 15
    e) 5.856, 19, 2.590, 18

29. What is the test statistic to determine if model 2 fits significantly better than model 1?
    a) 0.089 b) 1.098 **c) 3.840** d) 11.240 e) 22.698

30. What is the distribution of the test statistic comparing models 1 and 2?
    a) F(1, 18) b) F(4, 15) c) F(4, 19) **d) F(3, 15)** e) F(3, 19)