STATISTICS FOR ENGINEERS
ASSIGNMENT 15 ANSWERS

This practice assignment deals with all chapters, as does the final. The final will contain all tables you will need, but there will not be a page for formulas. You will, however, be allowed one page of notes, front only, standard paper size, hand-written.

Problem 15.1:

An engineer knows that the time until a lightbulb dies is distributed exponentially. Hypothetically, the expected lifetime for this shipment of lightbulbs is 50 days. What is the variance of the bulb lifetime? What is the probability that a bulb will last longer than 100 days? What is the probability that a bulb will die in less than one day? Given that a lightbulb has been burning for 45 days, what is the expected time from now until the lightbulb burns out?

Solution: Let T be the time (in days) until a lightbulb dies. We are given that $T \sim Exp(\lambda = 0.02)$.

- From out knowledge about the Exponential distribution, we know that the variance is the square of the mean, thus the variance is $\sigma_T^2 = 2500\text{d}^2$.
- To determine the probability that a bulb will last longer than 100 days, we calculate $\mathbb{P}[T > 100] = \exp(-0.02 \times 100) = \exp(-2) = 0.1353$.
- Similarly, $\mathbb{P}[T < 1] = 1 - \exp(-0.02 \times 1) = 0.01980$.
- One of the most important properties of the Exponential distribution is its "memoryless" property. Thus, we merely have to calculate the expected lifetime of the lightbulb, which is 50 days.

$\diamond$

Problem 15.2:

The boss asks you to determine if the number of defective parts produced per shift has changed since the machine was replaced. The number of parts produced each shift is Normally distributed, with mean 1500 and standard error 100. To perform the test, you select 100 parts at random in the shift just before the replacement and 100 parts at random in the shift right after the replacement. The number of defective parts in the pre-replacement sample is 15. The number of parts in the post-replacement sample is 10. Did replacing the machine change the rate of creating defective parts?

Solution: This is a good application of the proportions test. We have $p_1 = 0.150$, $p_2 = 0.100$, and $p = 0.125$. As the number of successes in each sample and the number of failures in each sample is larger than 10, we use the test on Page 428. The test statistic is

$$z = \frac{0.15 - 0.10}{\sqrt{0.125(0.875)(0.02)}} = 1.0690$$

This is a two-tailed test, and we find that $p = 0.2850 > \alpha$, therefore we cannot reject the null hypothesis that replacing the machine changed the defective rate. $\diamond$

Problem 15.3:

Your factory has four 6-hour shifts in the work day. No shifts overlap. The boss asks you to determine if the four shifts have the same productivity rate (the number of parts produced per shift). The relevent information is listed below, and a boxplot is provided in Figure 1. What should you tell your boss?

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Shift	3	838568	279523	1.561	0.2039
Residuals	96	17190649	179069		

Solution: You should tell your boss that you need the raw data, so you can run the non-parametric Kruskal-Wallis test. While the ANOVA procedure indicates no significant difference among the four shifts, the boxplot indicates a severe departure from Normality in the fourth shift. ◇

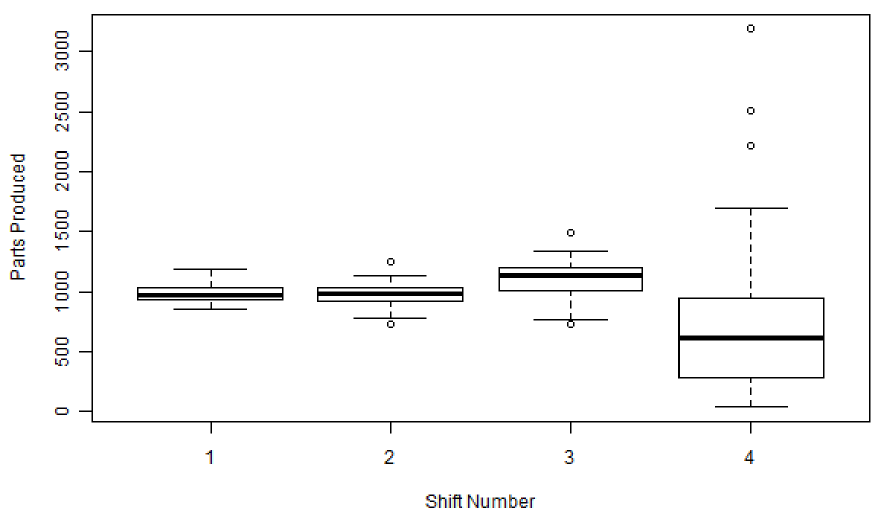

Figure 1. *Boxplot of the productivity of the four shifts at the factory.*

Problem 15.4:

Your research associate believes that there is a relationship between the age of the equipment and the number of bad parts it turns out per shift. To determine if she is correct, you collect data from your plant for a full month of shifts. In addition to the number of bad parts produced, you keep track of the day of the week, and the age of the operator. You then perform linear regression using the following statement.

```
glm(defectives ~ machine.age + doweek + operator.age)
```

The results are given in the table below.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-8328.5415	372.4920	-22.36	0.0000
Machine Age	62.6076	1.9451	32.19	0.0000
Operator Age	-2.4034	2.8196	-0.85	0.4028
Day = Monday	-19.4029	51.1912	-0.38	0.7081
Day = Tuesday	-32.5246	52.6201	-0.62	0.5426
Day = Wednesday	19.6915	51.2082	0.38	0.7041
Day = Thursday	-24.3602	50.6523	-0.48	0.6351

Do the results support your associate's statement? What should you do to improve the model? The linear model assumes that the residuals are distributed Normally. Will these residuals be distributed Normally? How do you know?

Solution: The results do, in fact, support your associate's statement. However, this is a bad model. The number of defectives is a count variable, which is bounded below by zero. As such, we should either transform the data using a log transform, or fit the data using an alternate distribution. Again, as the dependent variable is a count, the residuals will not — cannot — be distributed Normally. As such, a linear model is inappropriate. ◇

Problem 15.5:

An agricultural experiment was conducted to determine if any of four fertilizers were effective in reducing the height of dandelions in the field. There were four fertilizers and four fields. A two-way ANOVA was performed, and the raw computer output is in the following table.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Field	3	156.00	52.000	34.667	3.597e-10
Fertilizer	3	134.67	44.889	29.926	2.071e-09
Field:Fertilizer	9	0.00	0.000	6.34e-30	1
Residuals	32	48.00	1.500		

Are the effects of the fertilizer independent of the effects of the field?

Solution: According to the ANOVA table, the effects of the fertilizer *are* independent of the effects of the field. We know this because the interaction term is not statistically significant at the usual $\alpha = 0.05$ level $(F \approx 0; \nu_n = 9; \nu_d = 32; p = 1)$. ◇

Problem 15.6:

Your boss hands you a set of data containing the number of positioning tubes that have lengths outside allowable ranges (per day). This set contains 60 days' worth — 30 days' before and 30 days' after the machine change. You boss wants you to determine if changing the machine had a significant affect on the quality of the positioning tubes produced.

The three possible tests that spring to mind are the paired t-test, the t-test, and the Wilcoxon test. What are the assumptions of each of the three tests? What would you do to determine which test is most appropriate? Be specific, and remember that you have a computer to handle the actual testing, but you have to know which test is appropriate.

Solution: The two parametric tests assume Normality of the population. The paired t-test assumes that the measurements are repeated on individuals. The (independent) t-test assumes the measurements are independent. The Wilcoxon test (one sample) assumes that the population is symmetrically distributed. The Mann-Whitney (two sample Wilcoxon) test assumes the two populations are distributed identically except for their means (or medians). The Kruskal-Wallis (non-parametric ANOVA) makes the same assumption as the Mann-Whitney test.

To determine which test is most appropriate, I would check to see if the two samples are close to being Normally distributed. If they are, then I would use the (independent) t-test and not assume the variances are equal. If the samples obviously do not come from a Normal distribution, then I would use the Mann-Whitney test. ◇