Ph.D. Preliminary Examination
Statistics
May 26, 2017

NOTES:

1. The exam is worth 100 points.
2. Partial credit may be given for partial answers if possible.
3. There are 5 pages in this exam paper.

_ I have neither given nor received aid on this examination._

Name (print): ______________________________

Student ID : ______________________________

Signature, Date: ___________________________

=================================================================================================
1. **(10 points)** The continuous random variable $U$ has the following probability density function

$$f(u) = \frac{\theta^n u^{n-1} e^{-\theta u}}{(n-1)!}, \quad u>0,$$

where $n$ is a positive integer and $\theta > 0$. Calculate the mean and variance of random variable $U$.

2. **(20 points)** A manufacturer claims to control the variability in the length of life of the product so that the standard deviation $\sigma$ (in day) is no greater than 25. A wholesaler takes a random sample of 8 products and tests them in the laboratory, giving the following results in days.

$$725 \quad 741 \quad 706 \quad 711 \quad 735 \quad 697 \quad 745 \quad 752$$

(1) Test the hypothesis $H_0: \sigma = 25$ against the alternative $\sigma > 25$. Explain your results and state any assumptions you made.

(2) The manufacturer plans to change manufacturing process to reduce the variability in the length of life of the products. An experiment was conducted in which a number of samples using each process are produced and tested, with the following results.

| Process 1 | 724 | 743 | 705 | 711 | 736 | 699 | 745 | 752 | 740 | 705 |
| Process 2 | 725 | 740 | 715 | 732 | 720 | 702 | 740 | 741 | 738 | 725 |

Using an appropriate statistical test, investigate whether the manufacturer has been successful in reducing the variability using process 2. Explain your conclusions, stating any assumptions that you made.

3. **(20 points)** An experimental investigation wants to investigate the effect of two factors ($X_1$ and $X_2$) on the stress strength $Y$ (in Pa) by using a multiple linear regression model.
(1) Write down the regression model and explain the meanings and properties of the terms in the model.

The output of the regression model is as follows.

Regression Analysis: y versus xl and x2

Analysis of Variance Table

<table>
<thead>
<tr>
<th>Response: y</th>
<th>DF</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>2</td>
<td>2657.9</td>
<td>1328.95</td>
<td>229.53</td>
<td>4.404e-09</td>
</tr>
<tr>
<td>Residuals</td>
<td>10</td>
<td>57.9</td>
<td>5.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
<td>2715.8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coefficients:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|---------|
| Intercept| 52.57735   | 2.26617 | 23.00   | 5.45e-10 |
| xl       | 1.46831    | 0.12130 | 12.11   |         |
| x2       | 0.66225    | 0.04585 | 14.44   |         |

Residual standard error: 2.406 on 10 degrees of freedom
Multiple R-squared: 0.9787

Use the output to answer the following questions.
(2) Test the overall regression for significance at the 1% level, and explain the results.
(3) Write down the fitted regression equation of Y on X₁ and X₂ as defined above. Use it to predict the product lifetime with X₁ = 8 and X₂ = 35.
(4) Test regression parameters for statistical significance at the 0.1% level, quoting the critical value. What do the results imply about the effects of X₁ and X₂ on Y?
(5) Use the value of R² to comment on the overall fit of the model.

4. (30 points) A manufacturer wants to find a new material to improve product performance. The random variable Y indicates the number of material types before successfully finding one. Y has the distribution, parameter \( p(0 < p < 1) \), as follows,

\[ P(Y = y) = (1 - p)^y \cdot p \quad \text{for} \quad y = 0, 1, 2, \ldots. \]

This distribution has probability generating function

\[ \pi(t) = \frac{p}{1 - (1 - p)t} \quad \text{for} \quad t < (1 - p)^{-1}. \]

(1) Use the probability generating function to calculate the mean and variance
(2) Find the method of moments estimator of $p$.

(3) The random variables $Y_1, Y_2, ..., Y_n$ constitute a random sample from this distribution. Define $\bar{Y} = \sum Y_i / n$.

(a) Is $\bar{Y}$ is a biased or unbiased estimator of $1/p$? If biased, please find an unbiased estimator of $1/p$.

(b) Is $\bar{Y}$ (and the new estimator is needed) consistent? Why?

(c) The random variable $W$ is the number of random variables $Y_1, Y_2, ..., Y_n$ that take the value zero. (For example, if $n = 5$, $Y_1 = 1$, $Y_2 = 0$, $Y_3 = 3$, $Y_4 = 0$, and $Y_5 = 2$, then $W = 2$.) Calculate the distribution of $W$.

(d) Find an unbiased estimator of $p$ based on $W$ and give its variance.

5. (20 points) A tire company conducted a randomized block experiment on tire products, in which 5 treatments (A, B, C, D, E) were compared. The numbers of tire products per unit are shown as follows:

<table>
<thead>
<tr>
<th>Block</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>438</td>
<td>538</td>
<td>77</td>
<td>17</td>
<td>18</td>
<td>1088</td>
</tr>
<tr>
<td>2</td>
<td>442</td>
<td>422</td>
<td>61</td>
<td>31</td>
<td>26</td>
<td>982</td>
</tr>
<tr>
<td>3</td>
<td>319</td>
<td>377</td>
<td>157</td>
<td>87</td>
<td>77</td>
<td>1017</td>
</tr>
<tr>
<td>4</td>
<td>380</td>
<td>315</td>
<td>52</td>
<td>16</td>
<td>20</td>
<td>783</td>
</tr>
<tr>
<td>Total</td>
<td>1579</td>
<td>1652</td>
<td>347</td>
<td>151</td>
<td>141</td>
<td>3870</td>
</tr>
</tbody>
</table>

$\sum y = 3870 \quad \sum y^2 = 1395658$

(1) Calculate the means and standard deviations for the 5 treatment groups and examine the relationship between these two statistics.

(2) Construct the analysis of variance. State any assumptions required for the validity of this analysis, and explain how you would check whether these assumptions were reasonable.

(3) Calculate the standard error for a treatment difference. Hence identify pairs of treatments whose effects you would consider to be different.

(4) Explain why a square root transformation might be considered appropriate for these data. Transform the data for treatment A using a square root transformation and calculate the treatment mean.