Gx 2078

OLLSCOIL NA hÉIREANN, GAILLIMH

NATIONAL UNIVERSITY OF IRELAND, GALWAY

SEMESTER ONE EXAMINATIONS, 2002–2003

MA 337 – STATISTICS

Dr. D. Harrington,

Prof. J.P. Hinde

Time allowed: **Two** Hours.

Answer three Questions.

Relevant tables and formulæ are supplied.

Graph paper is available.

1. (a) The times between failures of a certain make of air-conditioning units can be described by an exponential random variable with mean 2000 hours, i.e. probability density function

$$f(x) = 0.0005e^{-0.0005x} \qquad x \geq 0.$$

What is the probability that a unit works for more than

i) 2000 hours;

ii) 4000 hours?

Explain the relationship between your answers.

(b) State the *Central Limit Theorem.*

For a random sample of size n from an exponential distribution with mean μ, use the Central Limit Theorem to show that $\overline{X}$, the sample mean, has an approximate $N(\mu, \frac{\mu^2}{n})$ distribution.

Use this approximation to show that

$$\left(\frac{\overline{x}}{1 + \frac{z_{\alpha/2}}{\sqrt{n}}}, \frac{\overline{x}}{1 - \frac{z_{\alpha/2}}{\sqrt{n}}} \right)$$

is an approximate $100(1-\alpha)\%$ confidence interval for μ, where $z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution.

(c) Data from a random sample of 100 air-conditioning units give a mean time between failure of 2360 hours. Assuming that the failure times are exponentially distributed, calculate an approximate 95% confidence interval for the true mean failure rate.

Are the data consistent with the manufacturer's claim that the mean time between failures is 2800 hours?

2. (a) State the relationships between the Normal, χ^2, t, and F distributions.
Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with mean μ and variance σ^2. What is $\mathrm{E}\left[\overline{X}\right]$, where $\overline{X} = (X_1 + X_2 + \cdots + X_n)/n$, the sample mean?

Prove that $\mathrm{E}\left[S^2\right] = \sigma^2$, where

$$S^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1}$$

Assuming that the population distribution for the $\{X_i\}$ is normal, explain why

$$\frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has a t distribution on $n-1$ degrees of freedom.
Hence derive a 95% confidence interval for the mean μ of a $N(\mu, \sigma^2)$ distribution, with σ^2 unknown, based on a random sample of size n. Explain why we use the term *confidence* rather than probability.

(b) As part of the maturation process, cheese is stored in underground caves. The following data are a random sample of "10kg wheels" of cheese giving the actual weight in kg:

9.8	8.6	8.4	8.8	8.7
10.2	10.4	10.1	10.5	

Calculate a 95% confidence interval for the population mean weight, stating any assumptions that you make. What do you notice about your interval?
Find a 95% confidence interval for the variance of the population, and hence obtain a 95% confidence interval for the population standard deviation.

3. (a) Explain the following terms used in hypothesis testing:

   - null hypothesis;
   - rejection (critical) region;
   - type I error;
   - power;
   - p-value.

   (b) Let X be a normal random variable with variance $\sigma^2 = 9$. Using a sample of size $n = 10$ and a significance level $\alpha = 0.05$, and writing the sample mean as $\overline{X}$, determine the rejection region for testing the hypothesis $H_0 : \mu = 12$ versus $H_1 : \mu > 12$.

   Show that the power function for this test is

$$1 - \Phi\left(\frac{13.56 - \mu}{\sqrt{0.9}}\right)$$

   where $\Phi(t) = P(Z \leq t)$ for $Z \sim N(0, 1)$. Plot this power function for $\mu = 12, 13, 14, 15$.

   Suppose that the sample size was increased to 100, determine the power function in this case and plot it for $\mu = 12, 12.5, 13, 13.5$.

   Comment on your plots.

   (c) Suppose that the observed sample mean is $\overline{x} = 12.8$, what would be your decision from the hypothesis test if the sample size was 10? Would your decision be the same if the sample size had been 100 and the same mean value was observed?

4. (a) The following frequency table gives the number of movements made by a foetal lamb in each of 240 consecutive 5-second intervals:

Number of movements	Number of intervals
0	182
1	41
2	12
3	2
4	2
5	0
6	0
7	1

Use a χ^2 goodness-of-fit test with $\alpha = 0.05$ to assess whether a Poisson distribution gives a suitable model for these data. (You will need to use some appropriate grouping.)

Comment on any aspects of the fit, or lack of fit, that you think are important.

(b) Explain how the χ^2 goodness-of-fit test can be used to test the normality of a sample of data from a continuous random variable. Be sure to refer to how the degrees of freedom should be calculated.

(c) The following data are on the lifetimes in hours of 9 light bulbs.

818 805 931 886 1136 948 1146 1592 810

Produce a normal quantile plot for these data by plotting the sorted data against the normal scores $(\Phi^{-1}(i/(n+1)),\ i = 1, \ldots, n$, where Φ is the cumulative distribution function of a standard normal distribution, $N(0, 1))$. Does the normal distribution provide a reasonable model here?

Explain what features we may be able to see in a normal quantile plot.

Formulæ

- Sample Mean

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

- Sample Standard Deviation

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}}$$

- Binomial Distribution

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, \ldots, n$$

$$\mathrm{E}[X] = np \qquad \mathrm{Var}[X] = np(1-p)$$

- Poisson Distribution

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x = 0, 1, \ldots$$

$$\mathrm{E}[X] = \lambda \qquad \mathrm{Var}[X] = \lambda$$

- Exponential Distribution

$$f(x) = \lambda e^{-\lambda x} \qquad x > 0$$

$$\mathrm{E}[X] = \frac{1}{\lambda} \qquad \mathrm{Var}[X] = \frac{1}{\lambda^2} = (\mathrm{E}[X])^2$$