

**Economics 573**  
**Problem Set 4**  
**Fall 2002**  
**Due: 20 September**

1. Ten students selected at random have the following "final averages" in physics and economics.

Students	1	2	3	4	5	6	7	8	9	10
Physics	66	70	50	80	60	70	55	90	75	85
Economics	75	70	65	88	60	85	60	97	82	90

- Put the data in a matrix whose rows are observations and whose columns are variables. Calculate the sample mean, sample variance, and sample correlation coefficient for these samples using MATLAB. You can do this using the `mean()`, `var()` and `corrcoef()` functions.
  - What is the critical value of a standard normal distribution such that 5% of the distribution lies in the upper tail? Use the `norminv` function in MATLAB to find this.
  - What is the critical value of a standard normal distribution such that 2.5% of the distribution lies in the upper tail? Use the `norminv` function in MATLAB to find this.
  - Suppose that the variance of the distribution of physics scores is known to be 169. Test the hypothesis that the mean final average in physics is 62 with  $\alpha = .05$  using the `ztest` function in MATLAB( `[h,sig,ci] = ztest(A(:,1),62,13,.05,0)` or `[h,sig,ci,zval] = ztest(A(:,1),62,13,.05,0)` ). Explain the output. The 0 as the last argument in `ztest` indicates a two-tailed test.
  - Suppose that the variance of the distribution of physics scores is known to be 169. Test the hypothesis that the mean final average in physics is less than 62 with  $\alpha = .05$  using the `ztest` function in MATLAB( `[h,sig,ci] = [h,sig,ci,zval] = ztest(A(:,1),62,13,.05,1)` ). Explain the output. The 1 as the last argument in `ztest` indicates a one-tailed test for the hypothesis that  $\mu \leq \mu_0$ .
  - Test the hypothesis that the mean final average in economics is 72 with  $\alpha = .05$  using the `ttest` function in MATLAB( `[h,sig,ci] = ttest(A(:,2),72,.05,0)` ). Explain the output.
  - Test the hypothesis that the mean final average in economics is 60 with  $\alpha = .05$  using the `ttest` function in MATLAB( `[h,sig,ci] = ttest(A(:,2),60,.05,0)` ). Explain the output.
  - Construct 95% confidence limits for the average economics score. Are the hypothesized mean scores for economics students in parts f and g included in this interval?
2. Use the same data as in problem 1.
- What is the appropriate test statistic to test the hypothesis that the variance of the economics scores is equal to 169?
  - What is the critical value of a chi-squared distribution such that 2.5% of the distribution lies in the upper tail? Use the `chi2inv` function in MATLAB to find this.
  - What is the critical value of a chi-squared distribution such that 2.5% of the distribution lies in the lower tail? Use the `chi2inv` function in MATLAB to find this.

- d. Can you reject the hypothesis that the variance of the economics scores is equal to 169 at a 5% level?
- e. Can you reject the hypothesis that the variance of the economics scores is greater than 190 at a 5% level?
- f. Test the hypothesis that the variance for students scores in physics is the same for students in economics. Use  $\alpha = .05$ .  $F(.025, 9, 9) = 0.248$  and  $F(.975, 9, 9) = 4.03$ .
3. Consider a random sample drawn from an unknown distribution. The data are as follows  $X = (18\ 29\ 19\ 14\ 20\ 11\ 15\ 33\ 21\ 14\ 19\ 15\ 14\ 18\ 16\ 21\ 33\ 18\ 17\ 27\ 27\ 25\ 20\ 11\ 25)$ .

- a. Using MATLAB, compute the sample mean and variance using the `mean()` and `var()` functions and also using the `sum()` function, i.e. `muhat=(1/n)*sum(x)`.

- b. Now compute the first four sample moments of the data. The  $r$ th sample moment is given by

$$\hat{\mu}'_r = \frac{1}{n} \sum_{i=1}^n x_i^r$$

- c. Now compute the first four central moments about the sample mean. The  $r$ th sample moment

$$\text{about the mean is given by } \hat{\mu}_r = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r$$

- d. Sample skewness is given by  $\alpha_3 = \frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}}$ . Find  $\alpha_3$  for this sample.

- e. Use the `skewness` function in MATLAB to calculate skewness for this sample.

- f. Sample kurtosis is given by  $\alpha_4 = \frac{\hat{\mu}_4}{\hat{\mu}_2^2}$ . Find  $\alpha_4$  for this sample.

- g. Use the `kurtosis` function in MATLAB to calculate kurtosis for this sample.

- h. Sort the data. Then obtain an estimate of the first quartile using the `median` function. The following code will be useful.

```
n=length(x);
xs=sort(x)
mid=round(n./2)
q1=median(xs(1:mid))
```

- i. Find the second and third quartiles.

- j. We can find percentiles of a given sample in general using the `interp1` function. The normal way to do this is to create an index for each observation equal to  $(i-0.5)/n$ . For example, if we have 10 observations, we index the tenth observation with the 95<sup>th</sup> percentile. After creating an index for each observation (say `phat`), we find the interpolated value of the series for other indices. If we wanted quintiles, we would let `p = [0.2, 0.4, 0.6, 0.8]` and then call the `interp1` function as follows.

```
qhat=interp1(phat,xs,p);
```

Use the `interp1` function to find the quartiles of the data for this problem.

k. Find the quartiles of the data using the  $y = \text{prctile}(x,[25\ 50\ 75])$  function.

4. Consider the same random sample as is problem 3.

a. Plot a histogram of the data using the following code where N gives the frequency counts and h the bin locations. The 9 indicates that the data will be allocated to 9 bins.

```
[N,h] = hist(xs,9);  
bar(h,N)
```

b. Use the normfit function to obtain maximum likelihood estimates of the parameters of a normal distribution based on the sample data. Also obtain confidence intervals for these estimates.

```
[muhat,sigmahat,muci,sigmaci] = normfit(x)
```

c. Verify that the estimates obtained are the same as we determined in the lectures by computing these estimates directly. You may use MATLAB to compute the estimates.

d. Use the expfit function to obtain maximum likelihood estimates of the parameters of an exponential distribution based on the sample data. Also obtain confidence intervals for these estimates.

```
[muhat,muci] = expfit(x)
```

5. Evaluate the following integrals using MATLAB.

a. 
$$\int_{-20}^{20} x \frac{1}{2\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-2}{2}\right)^2} dx$$

b. 
$$\int_{-20}^{20} x^2 \frac{1}{2\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-2}{2}\right)^2} dx$$

c. 
$$\int_{-20}^{20} (x - 2)^2 \frac{1}{2\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-2}{2}\right)^2} dx$$

6. Suppose  $X$  is a binary  $(0,1)$  variable which assumes a value of one with probability  $p$  and a value of zero with probability  $(1-p)$ . i.e.:

$$f(X) = p^x (1-p)^{1-x} \quad X = 0, 1$$

$$= 0 \quad \text{otherwise}$$

Note that  $E(X) = p$  and  $\text{var}(X) = p(1-p)$

- a. Show that  $\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$  is the MLE estimator of  $p$

$$\text{Hint: } L(X, p) = \prod_{i=1}^n p^{x_i} (1-p)^{(1-x_i)}$$

$$\begin{aligned} \log L &= \sum_{i=1}^n \log (p^{x_i} (1-p)^{1-x_i}) \\ &= \sum_{i=1}^n (x_i \log p + (1-x_i) \log (1-p)) \end{aligned}$$

- b. Show that  $\bar{X}$  is an unbiased estimator of  $p$ .

7. Consider the exponential density function given by

$$f(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} \quad 0 \leq x < \infty$$

- a. Show that the maximum likelihood estimates of  $\lambda$  is  $\bar{X} = \frac{\sum X_i}{n}$

- b. Given that  $E(X) = \lambda$ , show that  $\bar{X}$  is unbiased.

8. The density function for the normal distribution is given by

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (15)$$

Find the moment-generating function for  $X$  by integrating the following function.

$$M_X(t) = E[e^{tX}] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad (16)$$

You should obtain

$$\begin{aligned} M_X(t) &= \left[ e^{\frac{2\mu\sigma^2 t + \sigma^4 t^2}{2\sigma^2}} \right] \\ &= e^{\mu t + \frac{t^2 \sigma^2}{2}} \end{aligned} \quad (17)$$

as your answer.

9. The moments of X can be obtained from  $M_X(t)$  by differentiating. Find the first four moments of the normal distribution. For example, the first moment is

$$\begin{aligned} E(X) &= \frac{d}{dt} \left( e^{\mu t + \frac{t^2 \sigma^2}{2}} \right) \Big|_{t=0} \\ &= (\mu + t \sigma^2) \left( e^{\mu t + \frac{t^2 \sigma^2}{2}} \right) \Big|_{t=0} \\ &= \mu \end{aligned} \quad (18)$$

10. The joint pdf for two random variables X and Y is given by

$$F_{X,Y}(x,y) = 12xy(1-y) \quad (19)$$

for  $0 \leq x \leq 1$  and  $0 \leq y \leq 1$ .

- Show that this is a proper probability density function.
  - Find the marginal pdf of X.
  - Find the marginal pdf of Y.
  - Are X and Y independent?
11. The joint the two random variables X and Y are independent with marginal pdfs given by

$$\begin{aligned} F_X(x) &= 2x \\ F_Y(y) &= 3y^2 \end{aligned} \quad (20)$$

for  $0 \leq x \leq 1$  and  $0 \leq y \leq 1$ .

Find  $P(Y < X)$ .

12. The random variable Y has the following pdf

$$F_Y(y) = \theta y^{\theta-1} \quad (21)$$

for  $0 \leq y \leq 1$ .

- a. Find the first theoretical moment of Y.
- b. Assume you have the following random sample from this distribution:  $y = [0.42, 0.10, 0.65, 0.23]$ . Find the method of moments estimator of  $\theta$ .

13. The random variable X has the following pdf

$$F_X(x) = 3(1 - x)^2 \quad (22)$$

for  $0 \leq x \leq 1$ .

Consider a random variable  $Y = (1 - X)^3$ . What is the pdf for Y?

14. The random variable X has the following pdf

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (23)$$

for  $-\infty \leq x \leq \infty$ . Now consider the random variable  $Y = e^X$ .

- a. What are the bounds on Y?
- b. What is the inverse function  $X = g(Y)$ ?
- c. What is the pdf of Y?

15. Consider the following model

$$\begin{aligned}
 y_t &= \mu + \varepsilon_t \quad \text{or} \\
 y &= \mu \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \varepsilon \\
 &= \mu + \varepsilon \\
 E(\varepsilon_i) &= 0 \\
 E(\varepsilon_i^2) &= \sigma^2 \\
 E(\varepsilon_i \varepsilon_j) &= 0, i \neq j \\
 \Rightarrow E(\varepsilon \varepsilon') &= \sigma^2 I
 \end{aligned}$$

where  $\varepsilon$  is assumed to be distributed normally. Note the  $\mu$  is a single parameter and may not vary for each  $y$ .

- a. Derive the least squares estimator of  $\mu$ , call it  $\hat{\mu}$ . Do this by minimizing the following expression.

$$SSE = \sum_{i=1}^n (y_i - \mu)^2$$

- b. Write the answer to part a in matrix notation so that  $\hat{\mu} = a'y$  where  $a'$  is a row vector.
- c. The error vector can be written as  $e = y - i\hat{\mu}$  where  $i$  is a column vector of ones. Show that it can also be written as  $e = Ay$  where  $A$  is the matrix  $A = (I - ia')$ .
- d. Show that the matrix  $A$  is idempotent.
- e. Note that  $y$  can also be written as

$$y_i = \hat{\mu} + e_i$$

$$y = \hat{\mu} \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + e$$

$$\Rightarrow e = y - \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \hat{\mu}$$

Now show that  $e'e = \sum_{i=1}^n (y_i - \hat{\mu})^2$  can be written as  $y'Ay$ .

- f. Also show that  $e'e = \epsilon'A\epsilon$  by substituting  $\hat{\mu}i + \epsilon$  for  $y$  in  $e = Ay$ .

- g. How is the random variable  $\frac{\epsilon}{\sigma}$  distributed?

- h. Show how  $(e'e/\sigma^2)$  is distributed?

- i. What is the trace of the matrix  $A$ ?

- j. How is  $\frac{\sum_i (y_i - \bar{y})^2}{\sigma^2}$  distributed?