Economic Statistics I
Dr. Nisreen Salti
American University of Beirut
Fall 2006

Mock final: Solutions

1 Problem 1

The times that a cashier spends processing individual customer's order are independent random variables with mean 2.5 minutes and standard deviation 2 minutes. What is the approximate probability that it will take more than 4 hours to process the orders of 100 people?

ANSWER: The central limit theorem holds when $n = 100$. So we know that from a random sample of 100 people with processing times x_i where $i = 1, ..., 100$, the sample mean processing time $\overline{X} = \frac{\sum_{i=1}^{100} x_i}{100}$ has a normal distribution with mean 2.5 and standard deviation $\frac{2}{\sqrt{100}}$. So the sum of sample times $\sum_{i=1}^{100} x_i = 100\overline{X}$ has a normal distribution with mean $100 \times 2.5 = 250$ and standard deviation $100\frac{2}{\sqrt{100}} = 20$. So we can calculate the probability that the sum of processing times in the random sample exceeds 4 hours:

$$
\begin{aligned}
P\left(\sum_{i=1}^{100} x_i \geq 240\right) &= P\left(\frac{\sum_{i=1}^{100} x_i - 250}{20} \geq \frac{240 - 250}{20}\right) \\
&= P\left(Z \geq -0.5\right) \\
&= 1 - P\left(Z \leq -0.5\right) \\
&= 69.15\%
\end{aligned}
$$

2 Problem 2

Food poisoning outbreaks are often the result of contaminated salads. In one study carried out to assess the magnitude of that problem, the New York City Department of Health examined 220 tuna salads marketed by various outlets. A total of 179 were found to be unsatisfactory for health reasons. Find a 90% confidence interval for p, the true proportion of contaminated tuna salads marketed in New York City.

ANSWER: The confidence interval is:

$$\overline{p} \pm z_{0.95}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

where $z_{0.95} = 1.64$, and $\overline{p} = \frac{179}{220} = 81.4\%$.
So the confidence interval:

$$[77.09\%, 85.7\%]$$

3 Problem 3

Five independent samples, each of size n, are to be drawn from a normal distribution where σ is known. For each sample, the interval $\left(\overline{X} - 0.96\frac{\sigma}{\sqrt{n}}, \overline{X} + 1.06\frac{\sigma}{\sqrt{n}}\right)$ will be constructed. What is the probability that at least four of the intervals will contain the unknown μ?

ANSWER: The probability that at least 4 of the intervals will contain μ is the probability that exactly 4 and the probability that exactly 5 will contain μ. The probability that any such interval contains the true population mean μ is:

$$
\begin{aligned}
P\left(\overline{X} - 0.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + 1.06\frac{\sigma}{\sqrt{n}}\right) &= P\left(-0.96 \leq \frac{\mu - \overline{X}}{\sigma/\sqrt{n}} \leq 1.06\right) \\
&= P\left(-Z \leq 1.06\right) - P\left(-Z \leq -0.96\right) \\
&= P(Z \geq -1.06) - P(Z \geq 0.96) \\
&= P(Z \leq 1.06) - P(Z \leq -0.96) \\
&= 0.8554 - 0.1685 \\
&= 0.687
\end{aligned}
$$

The probability that exactly 4 of these intervals contain μ is:

$$
\begin{aligned}
\binom{5}{4}0.687^4(1-0.687)^1 &= \frac{5!}{4!1!}0.687^4 0.313^1 \\
&= 34.9\%
\end{aligned}
$$

The probability that exactly 5 of these intervals contain μ is:

$$\binom{5}{5}0.687^5 0.313^0 = 15.3\%$$

The probability that at least 4 of the intervals contain μ is:

$$34.9\% + 15.3\% = 50.2\%$$

4 Problem 4

According to a survey, 1 in every 410 Americans is a lawyer, but 1 in every 64 residents of Washington DC is a lawyer.

(a) If you select a random sample of 1500 Americans, use the Central Limit Theorem to calculate the probability that the sample contains at least one lawyer?

ANSWER: With a sample of 1500, the central limit theorem holds. The approximate probability that the sample proportion is larger than $\frac{1}{1500}$ is:

$$
\begin{aligned}
P\left(\overline{p} \geq \frac{1}{1500}\right) &= P\left(\frac{\overline{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \geq \frac{\frac{1}{1500} - \frac{1}{410}}{\sqrt{\frac{\frac{1}{410}\frac{409}{410}}{1500}}}\right) \\
&= P\left(Z \geq -1.39\right) \\
&= 0.9177
\end{aligned}
$$

(b) If the sample is selected from among the residents of Washington DC, what is the approximate probability that the sample contains more than 30 lawyers?

ANSWER: The probability that a sample from Washington DC of size 1500 contains more than 30 lawyers is:

$$\begin{aligned} P\left(\overline{p} \geq \frac{30}{1500}\right) &= P\left(\frac{\overline{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \geq \frac{\frac{30}{1500}-\frac{1}{64}}{\sqrt{\frac{\frac{1}{64}\frac{63}{64}}{1500}}}\right) \\ &= P(Z \geq 1.37) \\ &= 8.53\% \end{aligned}$$

(c) If you stand on a Washington DC street corner and interview the first 1000 persons who walk by and 30 say that they are lawyers, does this suggest that the density of lawyers passing the corner exceeds the density within the city? Explain.

ANSWER: If 30 people of the first 1000 that walk by are lawyers, the proportion is $\frac{30}{1000} = 3\% > \frac{1}{64} = 1.56\%$. This is an initial indication that this particular corner seems to be more densely frequented by lawyers than the average Washington DC corner. The probability that we observe a sample proportion of 3% or more when the population proportion is $\frac{1}{64}$ is:

$$\begin{aligned} P(\overline{p} \geq 3\%) &= P\left(\frac{\overline{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \geq \frac{3\%-\frac{1}{64}}{\sqrt{\frac{\frac{1}{64}\frac{63}{64}}{1500}}}\right) \\ &= P(Z \geq 4.49) \\ &\simeq 0 \end{aligned}$$

5 Problem 5

Research indicates that bicycle helmets save lives. A study reported in Public Health Reports (May-June 1992) was intended to identify ways of encouraging helmet use in children. One of the variables measured was the children's perception of the risk involved in bicycling. A four-point scale was used, with scores ranging from 1 (no risk) to 4 (very high risk). A sample of 797 children age 9-12 yielded the following results on the perception of risk variable: $\overline{X} = 3.39$, $s = 0.80$.

(a) Calculate a 90% confidence interval for the average perception of risk for all students age 9-12.

ANSWER:

$$\overline{X} \pm t_{\infty,0.95}\frac{s}{\sqrt{n}} = [3.34, 3.44]$$

(b) If the population mean perception of risk exceeds 2.5, the researchers will conclude that students in these grades exhibit an awareness of the risk involved with bicycling. Interpret the confidence interval constructed in part a in this context.

ANSWER: The 90% confidence interval doesn't contain 2.5, and we know that it is constructed to capture the true population mean with probability 90%. So we can be 90% sure that the true population mean exceeds 2.5, so that there is evidence that the students of these ages do exhibit an awareness of risk.

6 Problem 6

Greenpeace wants to test a randomly selected sample of n water specimens and estimate the mean daily rate of pollution produced by a mining operation. If Greenpeace wants a 95% confidence interval estimate with a bound on the error of 1 milligram per liter (mg/L), how many water specimens are required in the sample? Assume priori knowledge indicates that pollution readings in water samples taken during a day are approximately normally distributed with a standard deviation equal to $5mg/L$.

ANSWER: $\sigma = 5mg/L$. We are trying to figure out the appropriate sample size that gives a confidence interval of width 2. The width of a 95% confidence interval is $1.96\frac{5}{\sqrt{n}} \leq 1$.

$$n \geq 96$$

7 Problem 7

SAT scores, which have fallen slowly since the inception of the test, have now begun to rise. Originally, a score of 500 was intended to be average. The mean scores for 1991 were approximately 422 for the verbal test and 470 for the mathematics test. A random sample of the test scores of 20 baccalaureate students from urban secondary schools produced the means and standard deviations listed in this table:

	Verbal	Mathematics
Sample mean	419	445
Sample standard deviation	57	69

(a) Find a 90% confidence interval for the mean verbal SAT scores for students from urban schools, knowing that the population scores of students are normally distributed.

ANSWER:

$$\left[\overline{X} \pm t_{19,0.95}\frac{s}{\sqrt{n}}\right] = [396.96, 441.04]$$

(b) Does the interval you found in part (a) include the value 422, the true mean of the SAT score for 1991? What can you conclude?

ANSWER: The 90% confidence interval calculated includes the value 422, and since this confidence interval has a 90% chance of capturing the true population mean, we cannot rule out the possibility that the population mean verbal SAT score for students from urban schools is the same as for the general population of test takers.

(c) Construct a 90% confidence interval for the mean mathematics score for the urban secondary school baccalaureate students. Does the interval include 470, the true mean mathematics score for 1991? What can you conclude?

ANSWER:

$$[418.32, 471.78]$$

The interval above also includes 470 which means that we cannot, with 90% confidence, rule out that the true population mean for the mathematics score of urban students is the same as for the general population of test takers.

(d) Test at the 5% level of significance the null hypothesis that urban students do better on the verbal test than the overall population of students taking the test.

ANSWER: The hypothesis to test is specified in the question to be that urban students do better than the general population of test takers on average:

$H_0 : \mu > 422$

The rejection region is $\overline{X} < 422 - t_{19,0.95}\frac{s}{\sqrt{n}} = 399.96$.

Since we observe $\overline{X} = 419 > 399.96$ so we fail to reject the null hypothesis.

(f) Suppose you are told that in a test of the null hypothesis that urban students do worse on the math test than the overall population of test-takers, the decision rule is to reject the null hypothesis whenever the sample mean is above 490. What is the probability of Type I error in this case? What is the probability of Type II error if the true mean score for urban students is 475?

ANSWER: If the critical value is 490, and the null hypothesis is $H_0 : \mu < 470$, we know that the critical value is $470 + t_{19,1-\alpha}\frac{s}{\sqrt{n}} = 490$.So $t_{19,1-\alpha} = 1.296$. So $1 - \alpha$ is close to 90%, so $\alpha = 10\%$. This is the probability of Type I error. If the true mean score is 475, the probability of type II error is the probability of failing to reject H_0 when H_0 is in fact false. If the critical value is 490, the probability of failing to reject is the probability of observing a sample mean that is lower than 490 when $\mu = 475$:

$$\begin{aligned} P\left(\overline{X} < 490 | \mu = 475\right) &= P\left(t < \frac{490 - 475}{69/\sqrt{20}}\right) \\ &= P\left(t < 0.972\right) \\ &\simeq 0.85 \end{aligned}$$

8 Problem 8

A herbalist is experimenting with a juice extracted from berries and roots that may have the ability to affect the level of human body's calcium absorbtion. Assume that past experience suggest that human body's calcium absorbtion averages 0.5 with variance 0.09. You observe the calcium absorbtion level of 400 people after they took the juice for 6 months.

(a) Construct a test of the hypothesis that the juice does not affect the calcium absorbtion levels. Use a 10% significance level.

ANSWER: $H_0 : \mu \leq 0.5$

Critical value: $0.5 + z_{0.9}\sqrt{\frac{0.09}{400}} = 0.5192$

Decision rule: reject whenever $\overline{X} > 0.5192$

(b) Assume that the average calcium absorbtion level in the sample is 0.539. Test the null hypothesis.

ANSWER: given our decision rule, the observed $\overline{X}$ lies in the rejection region so we reject the null hypothesis that the juice doesn't affect calcium absorbtion.

(c) Assume now that the sample consisted of 40 individuals instead of 400 people, and that you want to be cautious because you do not want to claim to have a new drug when you really do not have one. What would you change, if any, from the test proposed in part a? Why? Do not compute the test again, just explain.

ANSWER: With a smaller sample size, we would need to make sure that the population is normally distributed or close to normally distributed for the central limit theorem to hold. We may also want to reduce the significance level of the test and increase the confidence level in order to make sure the results are more reliable.

9 Problem 9

The *placebo effect* describes the phenomenon of improvement in the condition of a patient taking a placebo –a pill that looks and tastes real but contains no medically active chemicals. Physicians at a clinic gave what they thought were drugs to 7,000 asthma, ulcer, and heartburn patients. Although the doctors later learned that the drugs were really placebos, 70% of the patients reported an improved condition. Use this information to test (at $\alpha = 0.05$) the placebo effect at the clinic. Assume that if the placebo is ineffective, the probability of a patient's condition improving is 0.5.

ANSWER: We are asked to test the placebo effect. And the hypothesis that gets tested is the null hypothesis. So if p is the true proportion of patients who do better after taking a placebo pill, the null hypothesis here is:

$H_0 : p > 0.5$

The critical value is: $0.5 - z_{0.95}\sqrt{\frac{0.5(1-0.5)}{7000}} = 0.49$.

The rejection region is $\widehat{p} < 0.49$. In our case $\widehat{p} = 0.7 > 0.49$ so our sample provides ample evidence at the 95% confidence level of the presence of a placebo effect.