Queen's University
Faculty of Arts and Sciences
Department of Economics
Economics 250 2008 Final

Instructions: 3 Hours

READ CAREFULLY. Calculators are permitted (no red stickers). At the end of the exam are several formulae and tables for the binomial, normal and t distributions. Answers are to be written in the examination booklet. Remember most of the grades are awarded for how you set up the problem and NOT for the calculation itself.

Please Note: Proctors are unable to respond to queries about the interpretation of exam questions. Do your best to answer exam questions as written.

You are to answer **ALL** questions. **SHOW ALL YOUR WORK**. There are a total of 100 possible marks to be obtained and marks are indicated for each question.

Answer all 7 questions.

1. (**20 marks**) Attendance at university classes has been declining in recent years. A professor wishes to determine whether there is a positive impact on being at class and calculates the sample mean and standard deviation for those regularly attending class and those who do not come to class. For those 25 attending class, the sample mean is 75 and standard deviation is 5. For those that do not attend class, the sample size is 30 with a mean of 71 and the standard deviation is 20.

   (a) What is the hypothesis test the professor should test and explain whether it should be one or two sided?

   $$\begin{aligned} H_0 &: \mu_c \leq \mu_{nc} \\ H_0 &: \mu_c > \mu_{nc} \quad \text{one sided since professor thinks} \\ & \quad \text{she teaching is likely improving the class} \end{aligned}$$

   (b) Do the formal hypothesis test at the 5% significance level and state the conclusion

   $$\begin{aligned} t &= \frac{\bar{X}_c - \bar{X}_{nc}}{\sqrt{\frac{s_c^2}{n_c} + \frac{s_{nc}^2}{n_{nc}}}} \sim t_{n_c + n_{nc} - 2} \quad \text{under } H_0 \\ t_{cal} \frac{75 - 71}{\sqrt{\frac{5^2}{25} + \frac{20^2}{30}}} &= 1.06 \\ t_{53,.05} &\approx F_z(0.95) = 1.65 \end{aligned}$$

   (c) Approximate the p-value as best as you can .

   $$p - value = P(Z > 1.06) = .5 - .3554 = .14$$

   (d) If the true population value is that students who attend class get on average 5 points higher, calculate the power of this test.

   $$\begin{aligned} & \text{Re}ject\ H_0 : \mu_c \leq \mu_{nc} \text{ if} \\ & \bar{X}_c - \bar{X}_{nc} > 1.645 \times \sqrt{\frac{5^2}{25} + \frac{20^2}{30}} = 6.2 \\ & P(\bar{X}_c - \bar{X}_{nc} - (\mu_c - \mu_{nc}) > 6.2 \mid \mu_c - \mu_{nc} = 5) \\ & Power = P\left(\frac{\bar{X}_c - \bar{X}_{nc} - (\mu_c - \mu_{nc})}{\sqrt{\frac{s_c^2}{n_c} + \frac{s_{nc}^2}{n_{nc}}}} > \frac{6.2 - 5}{\sqrt{\frac{5^2}{25} + \frac{20^2}{30}}} \right) \\ Power &= P(Z > .32) = 0.37 \end{aligned}$$

2. **(20 marks)** More information is preferred. Consider two independent populations surveying voter intentions. The first survey asks 100 people if they are going to vote and 40 say yes. In the second sample, 38 of the 120 people indicate they are going to vote.

   (a) What is the limiting distribution of the sample proportions for each assuming they both have the same population proportion p?

$$\begin{aligned}
\hat{p}_1 &= \frac{X_1}{n_1} \text{ is the fraction of intended voters population 1} \\
\hat{p}_2 &= \frac{X_2}{n_2} \text{ is the fraction of intended voters population 1} \\
\hat{p}_1 &\sim N\left(p, \frac{p(1-p)}{n_1}\right) \quad \text{asymptotically and} \\
\hat{p}_2 &\sim N\left(p, \frac{p(1-p)}{n_2}\right)
\end{aligned}$$

   (b) What is the best unbiased estimator that can be constructed using both sets of data and show its variance. Explain why this variance is smaller.

$$\begin{aligned}
\hat{p}_{pool} &= \frac{X_1 + X_2}{n_1 + n_2} = \frac{40 + 38}{100 + 120} = .35 \\
V[\hat{p}_{pool}] &= \frac{p(1-p)}{n_1 + n_2} \\
&\quad \text{clearly combining the two lowers variance} \\
V[\hat{p}_{pool}] &= \frac{p(1-p)}{n_1 + n_2} < \frac{p(1-p)}{n_1} \text{ or } \frac{p(1-p)}{n_2}
\end{aligned}$$

   (c) Construct the 99% confidence interval for this minimum variance estimator of the population mean and interpret?

$$\hat{p}_{pool} \pm Z_{\frac{.01}{2}} \times \sqrt{\frac{\hat{p}_{pool}(1-\hat{p})}{n_1 + n_2}}$$

$$.35 - 1.96 \times \sqrt{\frac{.35 \times (1 - .35)}{100 + 120}}$$

$(0.29, 0.41)$

If we do a large number of confidence intervals
95% of them are expected to bracket the true
population proportion

   (d) What null hypothesis range could you have where the null hypothesis would be retained?

Any hypothesis test with $p_0 \in (0.29, 0.41)$
will be retained at the 5% level of significance

(e) Knowing that more information is preferred, a cheap surveyor just uses the same population survey twice. That is he uses the first sample but doubling everything and says he has 200 people with 80 of them voting yes. Can this result in a lower variance just by pretending to have more data.? Why or why not?

$$
\begin{aligned}
& NO! \\
\hat{p}_{double} &= \frac{2 \times X_1}{2 \times n_1} = \hat{p}_1 = \frac{X_1}{n_1} \\
V[\,\hat{p}_{double}] &= V[\frac{X_1}{n_1}] = \frac{p \times (1-p)}{n_1}
\end{aligned}
$$

3. (**15 marks**) The Liberals, NDP and Bloc political parties of Canada intend to form a coalition government. Initially, there is equal probability of any party backing out of the coalition with probability 0.3. The coalition will fail if any group backs out. Assume on every matter other than the occurrence of a depression, the parties act independently. Suppose that if there is a depression the Liberals, Bloc, and NDP are likely to drop out of the coaltition with probability 0.2, 0.25, and 0.3 respectively. The probability of a depression has been calculated to be of equal probability over the interval over the next 100 days, so that at the end of 100 there is certain to have been a depression over the interval.

   (a) What is the unconditional probability that the Liberals drop out of the coalition?

$$P(L) = .3$$

   (b) What is probability that the Liberals drop out and there is a depression after 40 days?

$$
\begin{aligned}
P(L \cap D) &= P(D) \times P(L \mid D) \\
&= \frac{100-40}{100} \times .2 = .12
\end{aligned}
$$

   (c) What is the probability at the end of 100 days the government coalition still stands?

$$
\begin{aligned}
P(D) &= 1 \\
P(F) &= P(L \mid D) + P(N \mid D) + P(B \mid D) \\
-P(L &\mid D) \times P(N \mid D) - P(L \mid D) \times P(B \mid D) \\
-P(N &\mid D) \times P(B \mid D) + P(L \mid D) \times P(N \mid D) \times P(B \mid D) \\
&= .2 + .25 + .3 - .2 \times .25 - .2 \times .3 - .25 \times .3 + .2 \times .25 \times .3 \\
&= .58 P(S) = 1 - P(F) = .42
\end{aligned}
$$

4. **(10 marks)** There are three observations X_i $(i = 1, 2, 3)$ which are normally distributed with the same population mean 4 and sample variance of 100.. Each observation is correlated with the other two observations and all the correlation coefficients are 0.1.

(a) What is the sampling distribution for the sample mean?

$$\begin{aligned}
\bar{X}_3 &= \frac{1}{3}(X_1 + X_2 + X_3) \Longrightarrow E[\bar{X}_3] = \mu = 4 \\
V[\bar{X}_3] &= V[\frac{1}{3}(X_1 + X_2 + X_3)] \\
&= \frac{\sigma^2}{3} + \frac{2}{9}C[X_1, X_2] + \frac{2}{9}C[X_1, X_3] + \frac{2}{9}C[X_2, X_3] \\
&= \frac{\sigma^2}{3} + \frac{2}{9}(\sigma \times \sigma \times \rho + \sigma \times \sigma \times \rho + \sigma \times \sigma \times \rho) \\
&= \frac{\sigma^2}{3} + \frac{2}{9}(3 \times .1 \times \sigma^2) \\
&= \frac{100}{3} + \frac{2}{9} \times (3 \times .1 \times 100) = 40 \\
\bar{X}_3 &\sim N(4, 40)
\end{aligned}$$

(b) Calculate the probability that the sample means lies between 3.5 and 4.5

$$\begin{aligned}
& P(3.5 \leq \bar{X}_3 \leq 4.5) \\
&= P\left(\frac{3.5 - 4}{\sqrt{40}} \leq \frac{\bar{X}_3 - \mu}{\sqrt{V[\bar{X}_3]}} \leq \frac{4.5 - 4}{\sqrt{40}}\right) \\
P(-.08 &< Z < .08) = 2 \times .0319 = .06
\end{aligned}$$

5. **(15 marks)** Hypothesis tests and confidence intervals follow directly from the sampling distribution and this questions seeks to demonstrate your knowledge of that assertion. Suppose $X_i \sim NID(\mu, \sigma^2)$ for $n = 10$.

(a) What is the sampling distribution for the sample mean?

$$\bar{X}_{10} \sim N(\mu, \frac{\sigma^2}{n})$$

(b) What is the standardization transform for the sample mean?

$$Z = \frac{\bar{X}_{10} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0, 1)$$

(c) If the estimated sample mean is 10 and the estimated variance is 25, what is the 99% confidence interval. Interpret the interval. If we want to be more confident will we make a more or less precise statement?

$$\bar{X}_{10} \pm Z_{\frac{.01}{2}} \times \sqrt{\frac{s^2}{10}}$$

$$10 - 1.96 \times \sqrt{\frac{25}{10}}$$

$(6.9, 13.1)$

If we do a large number of them 99%

bracket the true population μ

More confident statement $\Longrightarrow$ less precison $\Rightarrow$ wider interval

(d) Suppose that we wish to test the null hypothesis $H_o : \mu = 9$ what would the sampling distribution of the sample mean be under this assumption?

$$\bar{X}_{10} \sim N(9, \frac{\sigma^2}{n})$$

$$\text{Under the Null} \Longrightarrow Z = \frac{\bar{X}_{10} - 9}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

(e) Illustrate that the test statistic for the null is the same as the standardizing transform for (d)

$$t = \frac{\bar{X}_{10} - 9}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

6. (**10 marks**) One sided hypothesis tests have the same level of significance as two-sided test but can have a different power

(a) Explain why this is true and use drawings to assist$\Longrightarrow$By foucusing on only one side of the distribution, leads to a larger crticial value (absolute value if on the left side) and hence there is an interval between t_α and $t_{\frac{\alpha}{2}}$ where the null would be rejected in the former and retained in the latter. If the null is false, power is increased over that interval

(b) For a 5% level of test of the population mean and $n = 25$ explain where one test can retain null and other could reject. What is the area of conflict of the one and sided test, give a number for conflict region

$$t_{.05,24} = 1.711 \quad t_{.025,24} = 2.064$$
$$(1.711, 2.064) \text{ interval of conflict}$$

(c) What happens to the size of the conflict region as n gets bigger. Try $n = 50$

$$Z_{.05} = 1.645 \quad Z_{.025} = 1.96$$
$$(1.645, 1.96) \text{ slight decrease in interval}$$

7. (**10 marks**) Suppose a gambler arrives in Los Vegas hotel to gamble with \$10,000. Suppose the amount of money an individual loses each day is independent from day to day and

$$L_d \sim N(d \times \$4000, \frac{1}{d} \times 2000) \quad d = 1, 2, 3 \ldots$$

If the gambler has less than \$100, the hotel asks the gambler to leave

(a) What is the probability the gambler is asked to leave after 1 day?

$$\begin{aligned} M &= 10000 - L_1 \\ P(M_1 &< 100) = P(L_1 > 9900) \\ &= P(\frac{L_1 - \mu_1}{\sqrt{\sigma_1^2}} > \frac{9900 - 4000}{\sqrt{2000}}) \\ P(Z &> 131.9) = 0 \end{aligned}$$

(b) What is the proability the gambler is asked to leave after 2 days

$$\begin{aligned} P(M_2 &< 100) = P(L_1 + L_2 > 9900) \\ E[L_1 + L_2] &= E[L_1] + E[L_2] = 4000 + 8000 = 12000 \\ V[L_1 + L_2] &= V[L_1] + V[L_2] = 2000 + 1000 = 3000 \\ P(L_1 + L_2 &> 9900) = P(\frac{L_1 + L_2 - (\mu_1 - \mu_2)}{\sqrt{V[L_1] + V[L_2]}} \geq \frac{9900 - 12000}{\sqrt{3000}} \\ P(Z &> -38.3) = 1 \end{aligned}$$