Queen's University
Faculty of Arts and Sciences
Department of Economics
Economics 250 2006 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 9 questions.

1. (**10 marks**) Assume each observation is from a normally and identically distributed population and consider the following estimator of the population mean μ

$$\ddot{X} = \frac{1}{4}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3 + 5$$

   (a) Define the bias of an estimator

   (b) What is the bias of this estimator $\ddot{X}$?

   (c) Consider the linear transformation of this estimator:

$$\tilde{X} = a + b\ddot{X}$$

   with a and b as constants. What values do these constants need to be to obtain an unbiased estimator of the population mean?

   (d) Calculate the variance of this estimator with your values from (c) and compare it to the minimum unbiased estimator of the population mean

   Answer

   A bias estimator is one in which $E[\hat{\theta}] \neq \theta$

   Bias of this estimator is $E[\ddot{X}] = E[\frac{1}{4}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3 + 5] = \frac{3}{4}\mu + 5$, an is clearly bias

   If we define

$$a = -5 \text{ and } b = \frac{4}{3} \Longrightarrow E[\tilde{X}] = \mu \text{ and will be unbiased}$$

$$V[\tilde{X}] = \frac{16}{9} \times V[\ddot{X}] = \frac{16}{9} \times \frac{3}{16}\sigma^2 = \frac{\sigma^2}{3}$$

   the minimum unbiased estimator is the sample mean

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ with } V[\bar{X}] = \frac{\sigma^2}{n}$$

   which is clearly smaller than $V[\tilde{X}]$

2. (15 **marks**) A city wishes to examine two kinds of road paint: "Lasts Really Long" and "Durable Road Paint". The city conducts road samples of 100 on "Lasts Really Long" and finds it last on average 450 days with a standard deviation of 30. The second product Durable Road Paint has a sample mean of 460 and a sample standard deviation of 10 on just 10 observations (This product is much more expensive)

(a) Provide the 99% confidence intervals for **each** population mean and interpret.

(b) Calculate the 99% confidence interval for the **difference** in population means assuming the variances are different

(c) Explain why the confidence interval for the difference in population means is not the same as the differences in the confidence intervals in (a)

(d) Would the hypothesis test that the two means are the same retained or rejected at the 1% level of significance?

(e) What is your advice to the city?

$$
\begin{aligned}
& \bar{X}_1 \pm t_{n_{1-1},\frac{.01}{2}} \times \frac{s}{\sqrt{n_1}} \\
& 450 - 2.576 \times \frac{30}{10} \\
& (442.3, 457.7) \text{ and} \\
& 460 \pm 3.25 \times \frac{10}{\sqrt{10}} \\
\text{since } t_{9,\frac{.01}{2}} &= 3.325 \\
& (449.7, 470.3)
\end{aligned}
$$

If we do a large number of these condidence intervals 99% will bracket the true value of μ_1 and μ_2

$$
\begin{aligned}
& \bar{X}_1 - \bar{X}_2 \pm t_{n_1+n_2-2,\frac{.01}{2}} \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \\
& 450 - 460 \pm 2.576 \times \sqrt{\frac{30^2}{100} + \frac{10^2}{10}} \\
& (-21.29, 1.3)
\end{aligned}
$$

The confidence interval of the population differences is NOT the difference of the individual population confidence intervals because of the standard errors are not linear.

The hpothesis test that the two means are the same would be retianed at the 1% level since the confidence interval brackets 0!

Since the "Durable Road Paint" is expensive and there is no statistical difference between the two, we would recommend going with "Lasts Really Long".

3. (**15 marks**) Suppose signal strength/quality in radio along the FM band for an existing station BOZO is a normally distributed variable with its center at 96 and a standard deviation of 12. students in 250 want to set up a radio station that is as close to BOZO's as possible. Canadian regulation requires that new stations can only overlap 30% of the transmission.

   (a) Draw a picture of the existing BOZO radio station band strength and draw another (lower side) to illustrate where the ECON 250 station is to go. Show what what we need to calculate.

   (b) What is the lower cutoff point that we can marks BOZO's territory?

   (c) Where does the Econ 250 station locate to obey the Canadian regulation?

   Answer

This is really like a power calculation!

$$\begin{aligned}
P(X_L &< c) = .30 \\
P\left(\frac{X_L-\mu}{s} < \frac{c-96}{12}\right) &= .3 \\
F_z(z_0) &= .7 \Longrightarrow z_0 = .53 \\
\frac{c-96}{12} &= -.53 \text{ (since we want the lower tail)} \\
c &= 96 + (12 \times -.53) \\
&= 89.6
\end{aligned}$$

Now no station's cut off point can exceed 89.6 on the dial.
To determine where the Econ 250 station can be by symmmetry (same standard deviation)

$$\begin{aligned}
P\left(\frac{X-\mu}{s} > 89.6\right) &= .3 \\
P\left(\frac{X-\mu}{s} > \frac{89.6-\mu}{12}\right) &= .3 \\
F_z(z_0) &= .7 \Longrightarrow z_0 = .53 \\
\frac{89.6-\mu}{12} &= .53 \text{ (since we want the upper tail)} \\
\mu &= 89.6 - (12 \times .53) \\
&= 83.24
\end{aligned}$$

4. **(15 marks)** One and two sided hypothesis tests seem mystifying to students when they first start to study inference and hypothesis tests. This question explores one and two-sided hypotheis tests by considering two independent samples where one group is given a parking ticket for illegal parking and the other group is given a warning. Of the 300 who first received a parking ticket 100 have returned to an illegal parking spot again. On the other hand of those 400 that received the warning 150 returned to an illegal parking spot.

   (a) Set up the null hypothesis and the alternative for both a one-sided and two-sided hypothesis test. Explain why you might prefer one over the other

   (b) Test your hypothesis for both one and two-sided alternatives at the 10% level of significance and make you conclusions

   (c) In general explain the circumstances in which a researcher might prefer a one-sided hypothesis test to a two-sided and discuss the consequences.

   Answer

   Two-Sided Hypothesis Tests

$$\begin{aligned}
H_o &: \pi_1 - \pi_2 = 0 \\
H_A &: \pi_1 - \pi_2 \neq 0
\end{aligned}$$

This is the two-sided hypothesis test (let π_1 and π_2 be the population proportion from those a parking ticket and warning respectively)

One Sided Hypothesis Tests

$$\begin{aligned} H_o &: \pi_1 - \pi_2 \geq 0 \\ H_A &: \pi_1 - \pi_2 < 0 \end{aligned}$$

The idea is that those that receive a warning are more likely to ignore parking signs since they paid no price the first time and might very well imagine that a warning would be given again. Hence the expected proportion of those returning from a warning is likely to be higher if the null is false.

Regardless of whether it is a one or two-sided alternative the test statistic is the same:

$$\begin{aligned} t &= \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim N(0,1) \text{ under the null} \\ p_1 &= .333 \quad p_2 = \frac{150}{400} = .375 \\ t_{cal} &= \frac{.333 - .375}{\sqrt{\frac{.333\times(1-.333)}{300} + \frac{.375\times(1-.375)}{400}}} \\ &= -1.15 \end{aligned}$$

For the two sided test

$$t_{n_1+n_2-2,\frac{.10}{2}} \sim Z_{\frac{.10}{2}} = \pm 1.645$$

for the one-sided alternative

$$t_{n_1+n_2-2,.1} \sim Z_{.10} = -1.282 \Longrightarrow \text{Reject } H_0 \text{ if } \mid t_{cal} \mid \leq Z_{.10} = -1.282$$

We can see that the null is retained at the 10% level regardless of whether we use a one or two-sided test

However, the one side alternative puts more probability in the tail where we think the null might be false. In this case there is the region (-1.645,-1.282) where test statistics would be rejected for the one-side alternative. If researchers have additional information then this can result in a test with more power without increasing the Type I eroor (rejecting a true null hypothesis)

5. **(15 marks)** Hypothesis testing at times seems to be a departure from some of the probability theory we learned. However, this is not true and we require some careful thinking to understand this. Let T be the event that the hypothesis we test is true.

   (a) In terms of the level of the test, α, make the conditional probability statement that links T and α. Explain in words

(b) What is the relationship between a $p - value$ and α

(c) What is the relationship then with $1 - \alpha$?

(d) Similarly describe probability of Type II error $1 - \beta$ and the corresponding power probability β

(e) If $P(T) = 1$ how many rejections of the null in n tests would we expect to see if the tests are based on independent data?.

(f) Suppose we have the situation where the validity of the null hypothesis is the outcome of some random experiment with $P(T) = \gamma$. What is the joint pprobability is true **and** we retain the null hypothesis?

(g) Do we in practice ever know anything about probability of T?
Answer

$$
\begin{aligned}
P(\text{Type I Error}) &= P(\text{Rejecting Null} \mid \text{Null Is True}) \\
&= P((\text{Rejecting Null}|T) \\
&= \alpha
\end{aligned}
$$

$$
\begin{aligned}
p - value &< \alpha \Longrightarrow \text{Reject Null at the } \alpha \text{ level} \\
p - value &\geq \alpha \Longrightarrow \text{Retain Null at the } \alpha \text{ level}
\end{aligned}
$$

$$
\begin{aligned}
1 - \alpha &= P(\text{Retaining Null} \mid \text{Null Is True}) \\
&= P((\text{Retaining Null}|T)
\end{aligned}
$$

$$
\begin{aligned}
1 - \beta &= P(\text{Retaining Null} \mid \text{Null Is False}) \\
&= P(\text{Retaining Null} \mid \bar{T})
\end{aligned}
$$

$$
\begin{aligned}
\beta &= P(\text{Rejecting Null} \mid \text{Null Is False}) \\
&= P(\text{Rejecting Null} \mid \bar{T}) \\
&= \text{Power}
\end{aligned}
$$

$$\text{Number of rejections} = n \times \alpha$$

$$
\begin{aligned}
P(T \cap \text{DO not Reject}) &= P(T) \times P(\text{Do not Reject} \mid T) \\
&= \gamma\alpha
\end{aligned}
$$

No we do not know whether the null is true or not. In fact it is either true or it is not, so the notion of probability of the null being true is problematic. If we new the null was true, why would we do hypothesis tests?

6. **(15 marks)** Distributions of random variables share some similar probability properties but can be completely different with respect to the range for them to have the same probability interval. Consider 4 random variables each with a different distribution:

$$\begin{aligned} Z &\sim N(0,1) \\ t &\sim t_8 \\ U &\sim Uniform(-3,3) \\ R &= -6+X \quad \text{where } X \text{ is the number of success with } n=20 \ \pi=.3 \end{aligned}$$

(a) Name two features of probability that all proper probability distributions have

(b) Draw a picture for each of the random variables above

(c) Explain whether it is continuous or discrete random variable, give its mean and for 3 out of the 4 give its variance (in one case you are not able to do so) .

(d) Suppose we want to leave .05 probability in the lower tail and .025 in the upper tail determine the upper and lower values for each random variable. For all answers, use a continuous random variable to calculate the probabilities. For any discrete random variable, use a continuous approximation (without continuity corrections) and explain what you are doing

Answer

Probability must be positive (greater than or equal to zero) for all discrete outcomes for discrete random variables and the positive (greater than or equal to zero). If we sum (integrate for continuous random variables) over all possible outcomes the probability must sum to 1.

All except R is a continuous random variable

$$\begin{aligned} P(Z_L &< Z<0)=.45 \text{ and } P(0<Z<Z_U)=.475 \Longrightarrow Z_L = \\ P(t_L &< t_8<0)=.45 \text{ and } P(0<t_8<t_U)=.475 \Longrightarrow t_L = \\ P(U_L &< U<0)=.45 \text{ and } P(0<U<U_U)=.475 \Longrightarrow U_L = \\ \text{Normal Approximation } R &\sim N(-6+n\times\pi, n\times\pi\times(1-\pi)) \Longrightarrow R\sim N(0,.42 \\ P(R_L &< R<0)=.45 \text{ and } P(0<R<R_U)=.475 \\ P\left(\frac{R_L-\mu_R}{\sqrt{n\times\pi\times(1-\pi)}}<Z<0\right) &= .45 \text{ and } P\left(0<Z<\frac{R_H-\mu_R}{\sqrt{n\times\pi\times(1-\pi)}}\right)=.475 \\ R_L &= \mu_R-1.645\times.42=-.69 \quad R_L=\mu_R+1.645\times.42 = \end{aligned}$$

7. (**15 marks**) Financial portfolios are combining two or more stocks in a single fund or holding with the aim of reducing risk. Consider a portfolio with the following two asset returns $A\sim N(5,2)$ and $B\sim N(5,3)$

(a) If we have 1 unit of both in our portfolio and we assume the stack have no correlation between the portfolios, what is the distribution, mean and variance of the combined portfolio

(b) If the stocks have a correlation of .3, what is the distribution, mean and variance of the portfolio

(c) If the stocks have a correlation of -.3, what is the distribution, mean variance of the portfolio

(d) Suppose you can hold any amount of the two stocks over the interval [0,2] (i.e. all in A or any combination along the real line up to all in B) what would provide the lowest risk (Hint; think about minimizing the variance of the portfolio)

Answer

$$R = A + B$$

Formula Sheet

Statistics Formulas

Notation

- All summations are for $i = 1, \ldots, n$ unless otherwise stated.
- $\sim$ means 'distributed as'

Population Mean

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Sample Mean

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

Population Variance

$$
\begin{aligned}
\sigma^2 &= \frac{1}{N}\sum_{i=1}^{N}[x_i \ - \ \mu]^2 \\
&= \frac{1}{N}\sum_{i=1}^{N} x_i{}^2 - \mu^2
\end{aligned}
$$

Sample Variance

$$s^2 \quad = \tfrac{1}{n-1}\textstyle\sum_{i=1}^{n}[x_i \ - \ \bar{X}]^2$$

Alternatively

$$s^2 = \frac{1}{n-1}[\sum_{i=1}^{n} x_i^2 \ - \ n\bar{X}^2]$$

Grouped Data (with k classes)

$$
\begin{aligned}
\bar{X} &= \frac{1}{n}\sum_{j=1}^{k}\nu_j f_j \quad \text{where } \nu_j \text{ is the class mark for class } j \\
s^2 &= \frac{1}{n-1}\sum_{j=1}^{k} f_j(\nu_j - \bar{X})^2
\end{aligned}
$$

Probability Theory

$$
\begin{aligned}
P(A \cup B) &= P(A) + P(B) - P(A \cap B) \quad \text{(additive law)} \\
P(A \cap B) &= P(B)P(A \mid B) \quad \text{(multiplicative law)}
\end{aligned}
$$

If the E_i are mutually exclusive and exhaustive events for $i = 1, \ldots, n$, then

$$
\begin{aligned}
P(A) &= \sum_{i}^{n} P(A \cap E_i) = \sum_{i}^{n} P(E_i)P(A \mid E_i) \\
P(E_i \quad | \quad A) &= \frac{P(E_i)P(A \mid E_i)}{P(A)} \quad \text{(Bayes' Theorem)}
\end{aligned}
$$

Counting Formulae

$$P_R^N = \frac{N!}{(N-R)!}$$

$$C_R^N = \binom{N}{R} = \frac{N!}{(N-R)!R!}$$

Random Variables

Let X be a discrete random variable, then:

$$\begin{aligned} \mu_x &= E[X] = \sum_x xP(X=x) \\ \sigma_x^2 &= V[X] = E[(x-\mu_x)^2] = \sum_x (x-\mu_x)^2 P(X=x) \\ \sigma_x^2 &= E[X^2] - (E[X])^2 \end{aligned}$$

The **covariance** of X and Y is

$$\begin{aligned} Cov[X,Y] &= \sigma_{xy} = E\left[(X-\mu_x)(Y-\mu_y)\right] \\ &= E[XY] - E[X] \times E[Y] \end{aligned}$$

The **correlation coefficient** of X and Y

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

If X and Y are independent random variables and a, b, and c are constants, then:

$$\begin{aligned} E[a+bX+cY] &= a + b\mu_x + c\mu_y \\ V[a+bX+cY] &= b^2\sigma_x^2 + c^2\sigma_y^2 \end{aligned}$$

If X and Y are correlated then

$$V[a+bX+cY] = b^2\sigma_x^2 + c^2\sigma_y^2 + 2bc\sigma_{xy}$$

Coefficient of Variation (CV)

$$CV = \frac{\sigma}{\mu} \times 100\% \text{ for population}$$
$$CV = \frac{s}{\bar{X}} \times 100\% \text{ for sample}$$

Univariate Probability Distributions

Binomial Distribution: For $x = 0, 1, 2, ..., n$ and :

$$Pr[X = x] = \binom{n}{x} \pi^x (1 - \pi)^{n-x}$$

$$E[X] = n\pi$$

$$V[X] = n\pi(1 - \pi)$$

Uniform Distribution: For $a < x < b$**:**

$$f(x) = \frac{1}{b - a}$$

$$E[X] = \frac{a + b}{2}$$

$$V[X] = \frac{(b - a)^2}{12}$$

Normal Distribution:For $-\infty < x < \infty$:

$$f(x) = (2\pi\sigma^2)^{-1/2} exp[\frac{-(x - \mu)^2}{2\sigma^2}]$$

$$E[X] = \mu$$

$$V[X] = \sigma^2$$

$$X \sim N(\mu_X, \sigma_X^2)$$
$$Z = \frac{X - \mu_X}{\sigma_X} \sim N(0, 1)$$

Estimators in General

If $\hat{\theta} \sim N(\theta, V[\hat{\theta}])$, say, then under appropriate conditions:

$$Z = \frac{\hat{\theta} - \theta}{SD[\hat{\theta}]} \sim N(0,1)$$

Confidence Intervals:

(a) $100(1-\alpha)\%$ confidence interval for θ : $\hat{\theta} \pm Z_{\alpha/2} SD[\hat{\theta}]$, with known variance.

(b) If $V[\hat{\theta}]$ is unknown, then

$$t = \frac{\hat{\theta} - \theta}{SE[\hat{\theta}]} \sim t$$

with appropriate degrees of freedom. $100(1-\alpha)\%$ confidence interval for θ : $\hat{\theta} \pm t_{\alpha/2} SE[\hat{\theta}]$, where $SE[\hat{\theta}]$ is an estimator of $SD[\hat{\theta}]$.

Estimating Means and Proportions

$$\bar{X} = \frac{1}{n}\sum X_i; \quad s^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2$$

$$E[\bar{X}] = \mu; \quad V[\bar{X}] = \frac{\sigma^2}{n}$$

$$p = \frac{X}{n}; \quad E[p] = \pi; \quad V[p] = \frac{\pi(1-\pi)}{n}$$

Differences of Means and Proportions for Independent Samples

$$s^2_{pool} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$E[\bar{X}_1 - \bar{X}_2] = \mu_1 - \mu_2; \quad V[\bar{X}_1 - \bar{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

If variances are assumed (i.e. $\sigma_1^2 = \sigma_2^2$) to be the same we may , estimate

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{s_{pool}^2 (\frac{1}{n_1} + \frac{1}{n_2})}$$

$$\begin{aligned} p_{pool} &= \frac{X_1 + X_2}{n_1 + n_2}; \\ E[p_1 - p_2] &= \pi_1 - \pi_2; \\ V[f_1 - f_2] &= \frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2} \end{aligned}$$