ECN 102: Analysis of Economics Data

Midterm Exam SQ 2025 – Answer Key

Remy Beauregard, Department of Economics, UC Davis

This exam consists of 5 short-answer questions (with sub-parts) and 5 multiple choice questions. You will have a maximum of 50 minutes to complete this exam. No additional time may be taken without prior accommodation. Show all work to receive full credit. You may use only the calculators provided by the instructor. For questions requiring computation, it suffices to express your final answer with 2 decimal places. There is a formula sheet and scratch paper provided at the end of this exam. You may remove these pages and discard them after the exam. If you plan to use any of these extra pages for your final answers, please write your name and student ID at the top of each scratch page to ensure they are not lost.

This exam is worth 70 points.

Question 1: Sample Means [10 points]

Suppose we know $J \sim (4, 4)$ and we draw 35 samples of size 49 and obtain 35 sample means.

a) How would we properly denote/write the sample mean of J? (You do not need to explain) [2 points]

$$\bar{J}$$

For the distribution of *sample means* of J:

b) What would you expect the population mean to be equal to? [2 points]

$$4$$

c) What would you expect the population standard deviation to be equal to? [3 points]

$$\frac{\sqrt{4}}{\sqrt{49}} = \frac{2}{7}$$

d) What range would you expect 99.7% of the sample means of J to fall under? Explain. [3 points]

$$4 \pm 3 \times \frac{2}{7} = (3.14, 4.86)$$

by 68-95-99.7 rule for normally distributed data by CLT with $n > 30$

Question 2: Summation Notation [6 points]

Please calculate the following sums:

a) [3 points]

$$\sum_{i=1}^{3} x_i \text{ with } x = [7, 3, 42, 5, 1]$$

$$7 + 3 + 42 = 52$$

b) [3 points]

$$\frac{1}{5} \sum_{i=3}^{5} (i + 1)$$

$$(4 + 5 + 6)/5 = 3$$

Question 3: Summary Statistics [10 points]

Suppose you obtain the following summary for a sample of data with `summarize x, detail`:

```
                            x
-------------------------------------------------------------
      Percentiles      Smallest
 1%     97.49081       97.43371
 5%     98.23293       97.54792
10%      98.6983       97.67799       Obs                 100
25%     99.28138       98.15409       Sum of wgt.         100

50%     99.80297                      Mean           99.85543
                       Largest       Std. dev.      .9636995
75%     100.4579       101.7433
90%      101.069       101.7931       Variance       .9287167
95%     101.5367       102.0256       Skewness        .094153
99%      102.231       102.4363       Kurtosis       3.142297
```

a) Interpret the mean, median, and skewness of this data. [3 points]

$\bar{x} \approx$ median; skew ≈ 0. Data is relatively symmetric

b) Interpret the kurtosis of this data. [3 points]

Kurtosis ≈ 3. Data has roughly normal distribution-width tails

c) Is this data (roughly) normally distributed? Explain. [4 points]

Yes; kurtosis is roughly 3 and skewness is roughly 0

Question 4: Univariate Inference [11 points]

You are given the following from Stata:

```
invttail(59,0.025)=2.0009954
invttail(59,0.05)=1.671093
invttail(58,0.025)=2.0017175
invttail(58,0.05)=1.6715528
invttail(57,0.025)=2.0024655
invttail(57,0.05)=1.6720289
```

Suppose you then use the `summarize` command to obtain the following results for the three variables below. The variables *price* and *weight* have been transformed to represent units of **thousands**.

```
(1978 automobile data)

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
         mpg |         59    21.45763    6.003798         12         41
      weight |         59    2.948475    .7801101       1.76       4.84
       price |         59    6.005695    2.585456      3.291     13.594
```

a) Can we say that *mpg* has higher dispersion than *price*? Why or why not? [3 points]

No, despite having a higher standard deviation, *price* has more dispersion than *mpg*.

$$CV_{mpg} = \frac{6.00}{21.46} = 0.28 < CV_{price} = \frac{2.59}{6.00} = 0.43$$

b) Provide a 95% CI for the mean of *weight*. Interpret your answer. [3 points]

$$\underbrace{2.95}_{\bar{x}} \pm \underbrace{\frac{0.78}{\sqrt{59}}}_{se} \times \underbrace{2.00}_{t^*_{58,0.025}}$$

This range has a $100(1-\alpha) = 95\%$ chance of including the true population value μ.

c) The claim is made that the average weight of a car in 1978 was below 3.25 thousand lbs. Using your sample data, test this claim at the 5% significance level. Clearly state your null and alternate hypotheses, your test statistic, and your conclusion. [5 points]

$$H_0 : \mu_{weight} \geq 3.25$$

$$H_A : \mu_{weight} < 3.25$$

$$t = \frac{2.95 - 3.25}{\frac{0.78}{\sqrt{59}}} = -2.95 < t^*_{58,0.05} = -1.67$$

Reject the null; we support the claim

Question 5: Bivariate Data [23 points]

a) Write down the population model for a bivariate regression of Y on X if linearity, unbiasedness, homoskedasticity, and independence hold. [2 points]

$$y = \beta_1 + \beta_2 x + u$$

b) Assuming our four population assumptions hold and our sample is large enough, complete the statement about the distribution of $\frac{b_2 - \beta_2}{\sigma_{b_2}} \sim$ ___(___, ___). [5 points]

$$\frac{b_2 - \beta_2}{\sigma_{b_2}} \sim N(0, 1)$$

Suppose we run the following regression and obtain these results:

```
      Source |       SS           df       MS      Number of obs   =        59
-------------+----------------------------------   F(1, 57)        =     15.39
       Model |  82.4431508         1  82.4431508   Prob > F        =    0.0002
    Residual |  305.262768        57  5.35548716   R-squared       =    0.2126
-------------+----------------------------------   Adj R-squared   =    0.1988
       Total |  387.705919        58  6.68458481   Root MSE        =    2.3142

------------------------------------------------------------------------------
       price | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         mpg |  -.1985808   .0506127    -3.92   0.000     -.299931   -.0972306
       _cons |   10.26677   1.127044     9.11   0.000     8.009901    12.52363
------------------------------------------------------------------------------
```

c) Write down the estimated (sample) relationship between *price* and *mpg*. [2 points]

$$\widehat{price} = 10.27 - 0.20 mpg$$

d) If we were to add `,robust` to the end of our regression code, list all the quantities in our regression results that would change. Why might we add this option? [4 points]

Our coefficients would **not** change. SEs and thus t-stats, p-values, and CIs **would** change. We would add this if we worried about heteroskedasticity.

Interpret the t-statistic and p-value for the OLS slope coefficient from the regression output:

e) What is the name of the test this referring to? What are its null and alternate hypotheses? [3 points]

Test of association of *mpg* on *price*;

$$H_0 : \beta_2 = 0$$
$$H_A : \beta_2 \neq 0$$

f) What is the conclusion of this test? [2 points]

Reject: p-val $\approx 0.000 < \alpha = 0.05$

g) How many degrees of freedom will this test have? [2 points]

$$n - 2 = 57$$

h) What would we expect the price to be for a car with an mpg of 25? [3 points]

$$10.27 - 0.2 \times 25 = 5.27$$

Multiple Choice [2 points each]

Only one answer is correct for each question. Choose the best possible answer.

MC 1

Remy wants to find out the most popular boba shop in Davis. For two days in a row, he goes to the quad at noon and asks everyone sitting there their favorite place for boba (after making sure they are a student). He then compiles this data into a single spreadsheet.

This data will be:

a) Experimental, categorical, panel

b) Experimental, numerical, time-series

c) Observational, numerical, cross-section

d) Observational, categorical, repeated cross-section

e) None of the above

MC 2

We define *residuals* as:

a) horizontal deviations of actual data points from the sample regression line

b) horizontal deviations of actual data points from the population regression line

c) vertical deviations of actual data points from the sample regression line

d) vertical deviations of actual data points from the population regression line

MC 3

We define *error terms* as:

a) horizontal deviations of actual data points from the sample regression line

b) horizontal deviations of actual data points from the population regression line

c) vertical deviations of actual data points from the sample regression line

d) vertical deviations of actual data points from the population regression line

MC 4

For a r.v. X with population mean μ and population variance σ^2, the *standard error* of X is:

a) An estimator for $\sigma_{\bar{X}}$

d) Always equal to $\sigma_{\bar{X}}$

e) An estimator for σ_X

f) Always equal to σ_X

g) None of the above

MC 5

Suppose a r.v. $Z \sim (6, 36)$. We would expect the transformed r.v. $\tilde{Z} \equiv (Z - 6)/6$ to be:

a) Normally distributed for any sample size

b) Normally distributed as $n \to \infty$

c) Distributed identically to Z

d) All of the above

e) None of the above