Econ 3040 A01 - Midterm - Fall 2022

Ryan T. Godwin

The exam is 70 minutes long, and consists of 72 marks (**approximately 1 mark per minute**). There are 9 short answer questions, each worth 4 marks. There is one long answer question with 9 parts, each part worth 4 marks. Write all answers in the provided exam booklet. You may only have a calculator and writing implements at your table. You may not use any books, notes, formula sheets, computers, or phones. A table of areas under the standard Normal curve is provided at the back of the exam.

DO NOT OPEN THIS EXAM BOOKLET UNTIL INSTRUCTED TO DO SO.

DON'T TOUCH! (Until instructed to do so).

Short Answer

1. A random variable Y is equal to 2 with probability 0.25, equal to 4 with probability 0.75. What is the expected value, and variance, of Y?

$$\mathrm{E}[Y] = 0.25(2) + 0.75(4) = 3.5$$
$$\mathrm{var}[Y] = 0.25(2-3.5)^2 + 0.75(4-3.5)^2 = 0.75$$

---

2. X and Y are two random variables, and their joint probability function is:

	$Y=-50$	$Y=2$
$X=-4$	0.70	0
$X=10$	0	0.30

What is the correlation between X and Y? Explain.

The correlation is 1. There is a perfect (positive) linear relationship between the two variables. If one of the variables is known, so is the other.

---

3. Explain why estimators are random variables.

Estimators are considered to be random variables because they are calculated from a random sample of data. For example the estimators $\bar{y} = \frac{1}{n}\sum y_i$ and $b_1 = \frac{\sum(y_i-\bar{y})(x_i-\bar{x})}{(x_i-\bar{x})^2}$ are random because the y_i are random.

---

4. How are the formulas for the least-squares estimators derived?

The least-squares estimators are derived by solving a calculus minimization problem where the sum of squared residuals are minimized.

---

5. What is the role of ϵ in the population model?

The error term ϵ contains all of the unobservable variables that influence y, besides the x variable(s). In any data set, there will always be "missing" variables that determine y. The error term ϵ accounts for these missing variables.

---

6. What is a least squares "predicted value", and what is a "residual"?

An LS predicted value is what we get when we "plug" the original x data into the estimated equation: $\hat{y}_i = b_0 + b_1 x_i$. A residual is the difference between the y value that is observed, and the y value that is predicted, for a particular x: $e_i = y_i - \hat{y}_i$.

---

7. What are some factors that might affect the precision (variability) of the least-squares slope estimator?

The LS estimator becomes more precise (has less variability) when:

- variance in ϵ is lower.
- variance in x is higher.
- the sample size n is larger.

---

8. Why is the least-squares estimator "good"? That is, why should we use the least-squares estimator instead of some other estimator?

The LS estimator is "good" in the sense that it gives the right answer on average (unbiased), has the lowest variance among other linear and unbiased estimators (efficient), and would give the "right" answer given an infinitely large sample size.

---

9. Describe a situation where R^2 would be equal to 1.

$R^2 = 1$ when x can be used to perfectly predict y. There would be no variation in ϵ. All the data points would lie on a straight line. All the predicted values will be equal to the actual values, and all the residuals are zero.

Long Answer

10. This question uses a dataset with $n = 2000$ and three variables: `wage` - the hourly wage of a worker in dollars, `education` - the number of years of education of the worker, and `economics` - a dummy variable taking on the value 1 if the worker has an economics degree, and 0 otherwise.

A least-squares model is estimated using R:

```
summary(lm(wage ~ education), data = mydata)
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.35165    0.98614   0.357    0.721
education    2.25278    0.06464  34.849   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.048 on 1998 degrees of freedom
Multiple R-squared:  0.378,     Adjusted R-squared:  0.3777
F-statistic:  1214 on 1 and 1998 DF,  p-value: < 2.2e-16
```

Use the above R code and output to answer questions (a) to (e).

a) What population model has been estimated?

$$wage = \beta_0 + \beta_1 education + \epsilon \tag{1}$$

b) What is the interpretation of the number 2.25278?

This is the "returns to education", or the marginal effect of education on wage. On average, when *education* increases by 1 year, *wage* increases by \$2.25 per hour.

c) How much of the variation in wage can be explained by years of education?

This is measured by the R^2: education explains 37.8% of the variation in wages.

d) Test the hypothesis that education has no effect on wage.

This hypothesis has already been tested by R. With a p-value of 0 (`<2e-16`), we reject the null that education has no effect on wage.

e) Use the above estimated model to predict how much a worker with 16 years of education will make.

We just "plug in" *education* = 16 into the estimated model: $\widehat{wage} = 0.35 + 2.25(16) = 36.35$.

Now, a model using the dummy variable is estimated (use this information to answer questions (f) to (i)):

```
summary(lm(wage ~ economics), data = mydata)
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  33.6273     0.2690 124.986  < 2e-16 ***
economics     3.6306     0.8551   4.246 2.28e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.42 on 1998 degrees of freedom
Multiple R-squared:  0.008942,  Adjusted R-squared:  0.008446
F-statistic: 18.03 on 1 and 1998 DF,  p-value: 2.277e-05
```

f) Is the dummy variable `economics` statistically significant?

A variable is "significant" if you would reject the null hypothesis that it has 0 effect on the y variable. R has already performed this test. With a p-value of 0.0000228 (`2.28e-05`), we reject the null. `economics` is significant.

---

g) What is the sample average wage for workers **without** an economics degree? What is the sample average wage for workers **with** an economics degree?

The sample average wage for workers without an economics degree is $b_0 = 33.63$ and the sample average wage for workers with an economics degree is $b_0 + b_1 = 33.63 + 3.63 = 37.26$.

---

h) An economics department is claiming that its economics graduates make \$4 per hour more than other graduates. Test this hypothesis using the above output.

The null and alternative hypotheses are:

$$H_0 : \beta_1 = 4$$
$$H_A : \beta_1 \neq 4$$

The t-test statistic for this hypothesis test is:

$$t = \frac{b_1 - 4}{s.e.(b_1)} = \frac{3.6306 - 4}{0.8551} = -0.43$$

Using Table 1, the p-value is $2 \times 0.3336 = 0.6672$. We fail to reject the null at any significance level.

---

i) Construct a 95% confidence interval around b_1.

$b_1 \pm 1.96 \times s.e.(b_1) = 3.6306 \pm 1.96 \times 0.8551 = [1.96, 5.31]$

END

Table 1: Area under the standard normal curve, to the right of z.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002