

ECO 307: Practice Exam 1

Exam Questions

Instructor: James M. Murray, PhD

Spring 2019

Section 1

Suppose a researcher working for local real estate agents seeks to understand the determinants of the sale price of homes. She collects data on sales price in dollars (`price`), number of bedrooms (`bdrms`), lot size in square feet (`lotsize`), size of the house in square feet (`sqrft`), and a dummy variable for whether or not the house is a colonial style house (`colonial`).

Suppose the researcher is interested in determining the how the \log of the sale price ($\log(\text{price})$) depends on all the above explanatory variables, and possibly interaction effects between colonial style houses (`colonial`) and size of the house (`sqrft`), and colonial style and the size of the lot `lotsize`.

Use such a regression to answer the questions below.

- A. (6 points) Is there statistical evidence that houses with more bedrooms have higher sales prices, accounting for the other explanatory variables? Test the appropriate hypothesis and interpret the result.
- B. (6 points) Report and interpret a 95% confidence interval for the impact one additional bedroom has on housing prices.
- C. (6 points) Is there statistical evidence that houses on larger lots have higher sales prices, accounting for the other explanatory variables? Test the appropriate hypothesis and interpret the result.
- D. (6 points) How much does a one standard deviation increase in lot size influence sales price for homes that are not colonial style, accounting for the other explanatory variables?
- E. (6 points) How much does a one standard deviation increase in square feet influence sales price for homes that are colonial style, accounting for the other explanatory variables?
- F. (6 points) Is there statistical evidence that colonial style homes have different average sales price than other style homes, accounting for the other variables in the model?

G. (7 points) Consider the R output below. Another researcher estimates a similar regression, but has different findings. The researcher sees the output below and claims there is strong statistical evidence that the number of bedrooms does lead to higher sales prices. In fact, based on the regression output below, he claims each additional bedroom leads to a 14% increase in house price.

Is the researcher correct? If so, explain why. If the researcher is wrong, explain why. If the research is wrong, explain whether he likely over-estimating or under-estimating the effect of bedrooms on house price.

```
lmhouse <- lm(log(price) ~ lotsize + bdrms + colonial + colonial:lotsize, data=data)
summary(lmhouse)
```

```
##
## Call:
## lm(formula = log(price) ~ lotsize + bdrms + colonial + colonial:lotsize,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.97035 -0.15253  0.00771  0.14336  0.60885
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.178e+01  1.335e-01  88.224 < 2e-16 ***
## lotsize       2.473e-05  7.791e-06   3.174  0.00211 **
## bdrms         1.452e-01  3.441e-02   4.219  6.22e-05 ***
## colonial     2.032e-01  9.639e-02   2.108  0.03802 *
## lotsize:colonial -1.897e-05  8.278e-06  -2.292  0.02446 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2545 on 83 degrees of freedom
## Multiple R-squared:  0.3293, Adjusted R-squared:  0.2969
## F-statistic: 10.19 on 4 and 83 DF,  p-value: 9.29e-07
```

```
confint(lmhouse)
```

```
##              2.5 %       97.5 %
## (Intercept)  1.151323e+01  1.204432e+01
## lotsize      9.233152e-06  4.022511e-05
## bdrms        7.675243e-02  2.136416e-01
## colonial     1.150100e-02  3.949458e-01
## lotsize:colonial -3.543617e-05 -2.505063e-06
```

Section 2

A workplace hold a number of “Wellness Wednesday” events which encourages and shares information about healthy behavior. The employer’s insurance provider is interested in getting higher participation in these events because they care about the employees as human beings, and they care even more about the money they will save on healthcare services it pays out as a result of unhealthy behaviors.

The employer surveys employees and measure the following variables:

- How many “Wellness Wednesday” events they attend in one year (**attend**),
- Sex (**male=1** for male, 0 otherwise)
- Age in years (**age**)
- How many happy hour events the employees attended during the same year (**drink**).

For each of the following research questions, state a appropriate univariate or bivariate hypothesis test that could help answer the question and state the null and alternative hypotheses.

- (6 points) Do men attend more or less “Wellness Wednesday” events on average than women?
- (6 points) Is the decision to participate in “Wellness Wednesday” events related to an employee’s age?
- (6 points) Do employees attend more happy hour events than “Wellness Wednesday” events?
- (6 points) Is participation in “Wellness Wednesday” events lower with people who attend more happy hour events?

Section 3

- (6 points) Define a sampling distribution.
- (6 points) Define multicollinearity. What problem may arise from multicollinearity? When may multicollinearity exist but not be problematic.
- (6 points) Suppose a researcher finds evidence for a positive interaction effect between a dummy variable on job sector (dummy variable equals 1 for manufacturing sector, 0 for all other sectors) and years experience (measured in years) on annual income. How would you explain this finding to a non-technical audience.