

Midterm Part 1

Week1

Please use the below information for Q1

The Baltimore data set has 506 observations, we create a model "**medv ~ rm + lstat + ptratio + tax + nox + indus**". The results of the summary:

```
call:
lm(formula = medv ~ rm + lstat + ptratio + tax + nox + indus,
    data = Boston)

Residuals:
    Min       1Q   Median       3Q      Max
-13.9802  -3.0470  -0.9347   1.7100  30.4545

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 19.145818   4.309359   4.443 1.09e-05 ***
rm           4.655928   0.431815  10.782 < 2e-16 ***
lstat       -0.545935   0.050641 -10.780 < 2e-16 ***
ptratio     -0.913819   0.131157  -6.967 1.03e-11 ***
tax         -0.002901   0.002225  -1.304  0.193
nox         -3.403117   3.478085  -0.978  0.328
indus        0.087187   0.061080   1.427  0.154
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.225 on 499 degrees of freedom
Multiple R-squared:  0.681,    Adjusted R-squared:  ???
F-statistic: 177.6 on 6 and 499 DF,  p-value: < 2.2e-16

Analysis of Variance Table

Response: medv
      Df Sum Sq Mean Sq F value    Pr(>F)
rm      1 20654.4  20654.4  756.4127 < 2.2e-16 ***
lstat    1  6622.6   6622.6  242.5339 < 2.2e-16 ***
ptratio  1  1711.3   1711.3   62.6727 1.593e-14 ***
tax      1    44.4    44.4    1.6246  0.2030
nox      1     2.4     2.4    0.0886  0.7661
indus    1    55.6    55.6    2.0375  0.1541
Residuals 499 13625.6    27.3
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Q1) what is the adjusted R-squared:

- A. 0.681
- B. 0.6772
- C. 0.6778
- D. 0.3228

Ans: B: $1 - \frac{((1 - 0.681) * (506 - 1))}{(506 - 6 - 1)}$ or $1 - \frac{[(13625.6/499)/(42716.3/505)]}{1}$ From wk1 Page 19 slide 1

Week2

For questions 2-4 we are going to use the *abalone* dataset. All questions are independent to each other. You do not have to use code for the following questions.

We will build a linear regression model using the Length (in mm) as the response variable and the Sex (Infant, Male, Female) and Height (in mm) as predicting variables. Using Sex = Female as the base case, we create 2 indicator (or dummy) variables SexInfant and SexMale for Sex = Infant and Sex = Male, respectively. Then, we fit the model and we obtain the output provided below. Use this output to answer questions 2 & 3.

Q2) For abalone with the same height, those that are male are expected to have greater length than female ones.

TRUE
FALSE

ANSWER

False (page 15 slide 1)

Q3) Choose the correct answer

- a. The estimate for the coefficient of the sex category infant is -0.045. This means that based on this model, if the abalone is infant instead of female, its length will decrease by 0.045 (on average), while holding height fixed.
- b. The estimate for the coefficient of the sex category infant is 0.199. This means that based on this model, if the abalone is infant instead of female, its length will increase by 0.199 (on average), while holding height fixed.
- c. The estimate for the coefficient of the sex category male is -0.004. This means that based on this model, if the abalone is male instead of female, its length will decrease by 0.240 (on average), while holding height fixed.
- d. The estimate for the coefficient of the sex category male is -0.004. This means that based on this model, if the abalone is male instead of infant, its length will decrease by 0.004 (on average), while holding height fixed.

ANSWER

a (Pages 10 & 11)

Q4) For question 4 we are adding 2 interaction terms $I_Height = Infant * Height$ and $M_Height = Male * Height$. Then, we run the regression and we obtain the following output:

	Estimate
Intercept	0.391
Height	1.193
SexInfant	-0.297
SexMale	-0.201
I_Height	1.898
M_Height	1.264

Choose the correct answer

- The slope for female abalone is 0.391.
- The slope for infant abalone is 1.193.
- The slope for male abalone is 2.457.
- The intercept for infant abalone is 0.391.

ANSWER

c (Page 29 slide 2)

Female slope = 1.193

Infant Slope = 1.193 + 1.898 = 3.091

Male Slope = 1.193 + 1.264 = 2.457

Infant Intercept = 0.391 - 0.297 = 0.094

Week3

Q5) Assume that $\text{Log}(y) = 16 + 0.72\text{Log}(x)$ is a model you are using in your research. How would you best explain the relationship between x and y?

- When x increases by 1%, y increases (on average) by 0.72%
- When x decreases by 1%, y increases (on average) by 0.72%
- When x increases by 1 unit, y increases (on average) by 0.72%
- When x decreases by 1 unit, y increases (on average) by 0.72%

Solution: A [Module 3 page 14]

Q6) Select the equation that can be used for a **log-linear model**:

- $Y = b_0 + b_1 * X$
- $Y = b_0 + b_1 * \log(X)$
- $\log(Y) = b_0 + b_1 * X$
- $\log(Y) = b_0 + b_1 * \log(X)$

Commented [PN1]: consider rewording to 'The relationship between x & y is almost equivalent to which of the following?'

I feel students might fight the wording here based on how slide 14 reads on the notes for mod3

Answer: C [Week 3, page 6, see below]

The Various Log Transformations

	Y	log(Y)
X	Model A: Level-level model $Y = b_0 + b_1 \cdot X$	Model C: Log-linear Model $\log(Y) = b_0 + b_1 \cdot X$
log(X)	Model B: Linear-Log Model $Y = b_0 + b_1 \cdot \log(X)$	Model D: Log-Log $\log(Y) = b_0 + b_1 \cdot \log(X)$

Note:

- The log() function in R computes the natural logarithm
- If a variable x has some values = 0, then use log(x+1) transformation

Georgia
Tech

6

Week4

Q7) Select the correct statement from below. Specificity of a model is defined as:

- Number of true positives divided by the sum of the number of true positives and false negatives.
- Number of true negatives divided by the sum of the number of true positives and false positives.
- Number of true positives divided by the sum of the number of true negatives and false positives.
- Number of true negatives divided by the sum of the number of true negatives and false positives.

Answer: d (Module 4, Page 26 Slide 1)

A study was conducted to measure the effect of two chemicals on the survival rate of beetles. The beetles were exposed to different levels of the chemicals. The output of a logistic regression model is provided below, where ptn represents the concentration of pyrethrin and pbt represents the concentration of piperonyl butoxide. Use this output to answer the next 2 questions.

```

Call:
glm(formula = Survived ~ ptn + pbt, family = "binomial", data = data)

Deviance Residuals:
Min 1Q Median 3Q Max
-9.5366 -2.4594 0.1223 3.9710 6.3566

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.58770    0.22958   15.63  <2e-16 ***
ptn          -4.32735    0.26518  -16.32  <2e-16 ***
pbt          -0.27483    0.01784  -15.40  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 718.76 on 11 degrees of freedom
Residual deviance: 299.43 on 9 degrees of freedom
AIC: 363.53

Number of Fisher Scoring iterations: 4

```

Q8) Interpret the coefficient for pbt.

- A 1-unit increase in the concentration of piperonyl butoxide (pbt) decreases the odds of a beetle surviving by 0.27483 when pyrethrin stays fixed.
- A 1-unit increase in the concentration of piperonyl butoxide (pbt) decreases the log odds of a beetle surviving by 0.27483 when pyrethrin stays fixed.
- A 1-unit increase in the concentration of piperonyl butoxide (pbt) decreases the number of beetles surviving by 0.27483 when pyrethrin stays fixed.
- A 1-unit increase in the concentration of piperonyl butoxide (pbt) decreases the probability of a beetle surviving by 0.27483 when pyrethrin stays fixed.

Answer: b (page 18, slide2)

Q9) What is the probability of survival for a beetle exposed to a pyrethrin (ptn) concentration of 0.5 and a piperonyl butoxide (pbt) concentration of 0.5?

- 0.011
- 0.382
- 0.784
- 0.216

Answer: c (page 19 slide 1)

Week5

Q10) Which of the following is true about a correlation between two quantitative variables?

- The greater the absolute value of a correlation coefficient, the weaker the linear relationship.

- B. Correlation of 0 indicates no linear relationship exists between two variables.
- C. The weakest linear relationship is indicated by a correlation coefficient equal to -1.
- D. A negative correlation means that if one variable gets bigger in positive side, the other variable tends to get bigger in positive as well.

Answer: B. Correlation of 0 means no linear relationship exists between two variables.

(Module 5 slide, page 2)

Q11) Which of the following is NOT an example of a natural experiment?

- A. Changing store policy to allow online orders to be picked up in all store locations
- B. Minimum wage is changed in one state, but not others
- C. A mobile carrier implements an unlimited data plan in some states but not others
- D. State emission laws are enacted in some cities but not in others

Answer: A. Changing store policy to allow online orders to be picked up in all store locations

(Module 5 slides, pg 16)

Q12) Which of the following is **NOT** a reason for selection bias?

- A. Self-selection
- B. Voluntary response bias
- C. Nonresponse bias
- D. Natural Experiment

Answer: D [Module 5, page 5]

Week6

Q13) Which measurement captures a stock's total risk?

- A. A stock's price
- B. The standard deviation of a stock's returns
- C. A stock's alpha
- D. The r-squared of a linear model regressing the stock's returns on the market's returns

Explanation: Standard Deviation is a measure of total risk

Q14) Which portfolio reduces firm-specific risk the best?

- A. A concentrated portfolio that contains stock from one company

Commented [SA2]: for Q10 option D I would say "A negative correlation means that if one variable gets larger in the positive direction, the other variable tends to get larger in the positive direction as well. "I think we had some students argue about the direction was not specified

- B. A diversified portfolio that contains stock from 60 companies over various industries
- C. Neither, both are exposed to the same level of firm-specific risk

Explanation: Diversifying your portfolio with 25-50+ stocks reduce firm-specific risk.

Q15) You can expect a risk free asset to have a beta close to which value?

- A. 2
- B. 1
- C. 0
- D. -1

Explanation: A risk free asset has a beta of zero

Week7

Q16) A portfolio had returns of 12% and a standard deviation of 6%. The risk-free rate of Treasury bills is 5%. What is the Sharpe ratio of the portfolio?

- A. 1.3
- B. .92
- C. 1.17
- D. 1.71

Ans. C. Week 7, Lesson 1 Slide 4 ; $(12-5)/6 = 1.16667$

Q: 17 Which of the following does **NOT** involve a type of transaction costs?

- A. A Stock with a bid price of \$25.15 and ask price of \$25.90.
- B. A stock at price \$60 with trade execution fee of \$3.
- C. A stock planned to be bought at \$15 changes to a price of \$15.85 at the time of submission of buy order.
- D. A stock bought directly from the seller.

Ans D. does not mention any transaction costs while other options do. Week 7, Lesson 1 Slide 9

Q 18. Which of the following is true about Bid and offer prices?

- A. Buyers of stock would ideally look for the lowest bid price
- B. Sellers of stock would ideally look for the highest bid price
- C. Sellers of stock would ideally look for the lowest ask price
- D. Buyers of stock would ideally look for the highest ask price

Ans B. Seller would want to sell their stock at the best bid price that is the highest bid price
Week 7, Lesson 2 Slide 9;

Week8 (2 Questions)

Q19) Suppose we ran a factor regression for a stock fund to see which factors explains its returns and got the following output:

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	-0.003	0.004	-0.724	0.473
Mkt-RF	0.757	0.140	5.394	1.634E-06
SMB	-0.721	0.159	-4.543	3.238E-05
HML	-0.056	0.165	-0.338	0.736

Where Mkt-RF is the excess market return, SMB is the Size factor and HML is the Value factor. Looking at the coefficients of the factors and their corresponding significance, this fund is most likely a:

- A. Growth Fund
- B. Large-Cap Fund
- C. Small-Cap Fund
- D. Value Fund

Answer: B (Week 8 Lesson 2 Slide 10,12)

Explanation: As the coefficient of the Size factor is negative and significant, this fund most likely is a Large-Cap Fund. Please note although the coefficient of Value factor is negative, it is not significant. Hence, this fund in all probability will not be a Growth Fund.

Q20) In recent times, one of the most popular strategies on Wall Street has been to invest in high growth (or equivalently low value) stocks. People who jumped on the strategies early with names like Shopify, Netflix, Tesla, and Beyond Meat have drastically outperformed the broad market by betting on the growth factor (or equivalently against the value factor). Famed value investors such as Warren Buffett and David Einhorn who made their fortunes betting on the value factor prior to 2008 are now experiencing some of the worst returns in their investing careers.

Given your understanding of the long run behaviour of the value factor which of the following statements is **NOT** true.

- A. Although the value factor has underperformed in recent years, it is normal for systemic factors such as value to experience times of cyclical fluctuation.
- B. As more investors put money into high growth stocks the high growth premiums presently observed are more likely to be crowded out and diminish in the long run
- C. The value factor as we know it no longer exists
- D. As more investors flee high value stocks (due to the presently negative returns) the future return premium for the value factor could likely increase

Answers: C (Week 8 Lesson 2 Slide 13 & 14); Answer A is true it is consistent with our knowledge of long-term factor behaviour in that systemic factors such as value will go through periods of both outperformance and underperformance due to cyclical factors. Answer B is true and is consistent with the principals of arbitrage that as more people join in on a profitable trade the profit will eventually disappear as competition for the assets push prices up wiping out any excess return premium. Answer C is false and is the correct answer choice as its recent underperformance of value is not sufficient-enough a time period or persistence to warrant a

rejection of the factor especially considering it allows for cyclical fluctuations. Answer D is also true as the inverse case of answer B. As investors continue to sell value stocks, eventually they will become so oversold in comparison to high growth stocks that a positive premium will eventually likely emerge.