

**Due: Feb 2, 4pm. (same day as the midterm).**

1. (8 marks) In the article “Randomness in the Stock Market”, it is reported that “A set of randomly chosen stocks typically equals or outperforms the advice of the majority of investment newsletters”. This is true even though most newsletter authors are highly-educated and respected in the investment business. Why do think this is so? (Hint: Think of a reason mentioned often in this course.)

A1. The symmetric random walk simulation showed that spurious trends seem to result even when there is no real predictability of the time series. It is likely that many people believe these trends are predictable, since it is counter-intuitive to realize this fact about time series that are close to symmetric random walks. (Even if the experts know this, the clients they serve likely do not).

2. (7 marks) Zipf’s Law is an example of a model that is “wrong” but useful, when applied to the populations of a nation’s cities. Explain.

A2. The model worked for Canada, and for the USA, but was very wrong for Australia. The useful outcome was that the urbanization of Australia was different from Canada and the USA, and so the analyses provide a starting position for a further look at the reason why.

3. (7 marks) In the article “Predicting Quality and Prices of Wines”, why was it necessary to adjust for “AGE” before observing the influence of weather on the relative prices of vintage wines?

A3. The prices of wines will increase with AGE even if there were identical weather conditions in the years analyzed. But the weather variables do not have a systematic increase with AGE. So to best see the relationship of price to weather, you need to look at the AGE-adjusted prices, in comparison with the weather data.

4. (8 marks) In the example of the five years of gasoline consumption that was used to illustrate the power of smoothing of a time series, it was mentioned that a moving average method could have been used for the smoothing. What determined the appropriate order of the moving average used for this purpose? Hint: The “order” of the moving average was the number of data values averaged at each step.

A4. The context of the data. The reasonability of the pattern provided by the moving average, noting for example that the annual-pattern might well be seasonal with period 12 months, is what determines how much smoothing is appropriate. Too little smoothing provides a chaotic pattern that is not credible, while too little smoothing decreases the amplitude of the seasonal pattern without changing the seasonal pattern otherwise.

5. (7 marks) In the Turkey mail article (Advertising as an Engineering Science”), why were no small p-values mentioned to support the claimed influences of SUBJECT and DAY-OF-WEEK on CLICK-THROUGH\_RATE?

A5. The number of emails sent out is huge, so relationships noted in simple graphs are unlikely to be merely effects of randomness. However, there is reference to a logistic regression analysis in which proper hypothesis testing was done, but it was deemed to be too complex for the intended audience of the article (i.e. for beginning students of statistics). (This discussion on p 386.)

6. (8 marks) In your Assignment #4, question 6. asked you to explain “why averaging can make a portfolio of risky investments into a relatively stable portfolio investment”. What are the requirements of a portfolio that make this a valid claim?

A6. A positive average return and independence of the investment outcomes of each company.

7. (8 marks) In the article “Advertising as an Engineering Science”, there is frequent reference to the fact that the study described was an “experiment”. What feature of the study made it a true experiment, **and** why was this study design chosen?

A7. The investigator assigned the features of the comparison groups to the experimental units – this is the feature (SUBJECT and DAY-OF-WEEK were assigned at random to the registrants.) This design was chosen because the investigator wanted unambiguous information about how these two variables affected (or “caused” )the CLICK-THROUGH-RATE.

8. (7 marks) How would you use tosses of a fair coin to simulate equal-probability selection of the integers {1,2,3,4,5,6,7,8}? If you used this procedure to create a sample of 25 values, estimate how likely would it be to get a sample mean less than 4.0? Hint: The SD of {1,2,3,4,5,6,7,8} is 2.5.

A8. Mean is obviously 4.5, and SD is given as 2.5. So 4.0 is 0.5 below the mean. The sample mean has  $SD = 2.5/\sqrt{25} = 0.5$ . The normal distribution approximates the probabilities for the sample mean and the probability for the sample mean to be less than 4.0 is about 16%. (68% within 1 SD -> 32% outside of  $\pm 1SD$  -> 16% below - 1SD).

KLW 2010/01/29