Business Analyst Take-Home Project

*Copyright Applied Underwriters, Inc. All Rights Reserved.*

Currently, patients with a certain category of back injuries usually take a lifelong prescription of powerful pain medication. This medication is known to decrease life expectancy. Trials of a new surgical procedure have been conducted on some patients. If successful, the surgery immediately rids the patient of the ongoing need for medication. The surgery can only be attempted once. If the surgery is unsuccessful then there are no new adverse effects but the patient must continue on medication as before.

You have separately been provided with credentials to connect to a VPN and a remote machine. You may use any software on the machine. ***Sufficient software for this project is available there***, as are the source data tables in FoxPro (.dbf) format. The source data tables are in a directory on the desktop of the test machine. Your answers to Part I should be loaded into the provided tables QuestionA and QuestionB on the local SQL Server database Analyst\_Project. Any files used (including all of your code) should be stored in the directory on the desktop named Results, along with your answers for Part II. Assume today’s date is **January 1, 2018**.

Be prepared to talk about your results in a short 20 minute technical interview. Any experience you have with R, SQL Server, Excel or FoxPro is likely to be an advantage. However, we don’t require you to have prior experience, and this project has been completed successfully by candidates with no prior knowledge of SQL, R or FoxPro.

Good Luck!

# Part I

## Source Data

* LifeTableMale: a life table for males. Assume probability of death **within** each age-year is *uniformly distributed*.
  + Age: the age-year, ranges from 0 to 119. You may assume death is immediate for individuals who reach 120.
  + Prob: the probability of dying within the corresponding age-year.
  + MedProb: the probability of dying within the corresponding age-year for an individual on the medication.
* LifeTableFemale: a similar life table for females.
* Medication:
  + Year: the calendar year.
  + Month: the calendar month.
  + Cost: the cost (in dollars) of one patient’s supply of pain medication for that month.
* Patients: a sample of 600 patients.
  + ID: the patient’s identification number.
  + Gender: the gender of the patient.
  + DoB: (Date of Birth) the patient’s date of birth.
  + DoM: (Date of Medication) the date on which the injury occurred and the patient began taking the medication.

## Questions

1. (10 points) For each patient, calculate their life expectancy if **they do not take the medication from now on**. Then calculate what their life expectancy would be if they take the medication for the rest of their life. For each patient, how much higher is the life expectancy without medication?
2. (5 points) What is the total expected future cost of the medication for each patient, assuming they take it for the rest of their life? Hint: this is ***not*** equivalent to the cost of the medication over the life expectancy.

**These are not equivalent because the scaling in price is not linear (so expectation of the nonlinear function is not a linear function of the simple expectation).**

**Try working with monthly probability of dying and attach price.**

**Check if R’s LifeTable assumes the midpoint (0.5x); if so, might want to model with each 6 mo or stagger 1 year.**

# Part II

## Source Data

* PatientsSurgery: surgical trial results for 200 patients from the Patients table.
  + ID: the patient’s identification number.
  + DoS: (Date of Surgery) the date on which the patient received surgery.
  + Surgery: whether the surgery was a success or failure. Success = 1.

## Questions

Consider gender and the age at which the surgery was performed as the pool of variables in **predicting the probability of a successful surgery**.

1. (6 points) Explain why an ordinary least squares method would not be a good model choice. In particular, explain which assumptions of OLS are violated and why.
2. (10 points) Fit an appropriate model; use Likelihood Ratio Tests for any model selection.

~~A – In our dataset, we’re only looking at patients who are currently alive on January 1, 2018 and not those who have died either medicated or unmedicated. Hence the symmetry assumption for an ordinary least squares (OLS) method. (this has nothing to do with predicting success/failure via age\_at\_surgery and gender.~~

In OLS regression, we must have a linear relationship between predictor and predicted variables; however, logistic regression does not make this assumption.

OLS linear regression assumes that the residuals are normally distributed; but look at how the they’re distributed in our plot (sawtooth);

Should not assume that regression is appropriate!

Consider opioid prescription for back pain

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5582019/>

“Using one of the nation’s largest commercial insurance databases, we conducted a retrospective cohort study of 8,377 adults, aged 21–63 years, who underwent lumbar spinal fusion surgery between January 1, 2009 and December 31, 2012. Long-term opioid use was defined as ≥365 days of filled opioid prescriptions in the 24 months following lumbar fusion. Multivariable logistic regression was used to calculate adjusted odds ratios (ORs) and 95% confidence intervals for the risk of long term opioid use following lumbar fusion.”

Reference ~ <http://www.lifeexpectancy.org/lifetable.shtml>