Quantitative Reasoning and Statistical Methods
for Planners
Test-Out Exam

Ezra Glenn (eglenn@mit.edu; Room 7-337)

Sample Test

Name: ______________________  email: ______________________
Problems

* For all problems involving questions of statistical significance, please use a 95% confidence level (i.e., $\alpha = .05$).

1. Before it was used as the name of a popular search engine, the word “google” was used to describe a specific large number, represented as a “1” followed by 100 zeros.

   How many zeros do you need to write $\frac{1}{2}$ of a google?

2. In 2008, 412 cadets took the police academy officer training exam in Pawtuckaway County. The following numbers represent a sample of five of these scores (the test is graded between 0 and 65): 48 32 52 40 59

   (a) What is the mean of these scores?

   (b) What is the standard deviation of this sample?
(c) If these scores represent a simple random sample of all 412 cadets taking the test in that year, what would be your best estimate of the mean score for the entire class? Please give this estimate with a 95% confidence interval.

(d) Assume you were to examine the scores of 45 more cadets (for a total of 50). Surprisingly, the mean and standard deviation for this larger sample remain the same as those found in items 2.a and 2.b (above). How would you revise your estimate and confidence interval, if at all? Why or why not?
3. For the following questions, please consult the histograms A-D.

(a) Rank the distributions, from the one with the lowest mean to the one with the highest mean.

(b) Rank the distributions, from the one with the lowest median to the one with the highest median.

(c) Rank the distributions, from the one with the lowest standard deviation to the one with the highest standard deviation.

Name:
4. Since the September 11th terrorist attacks, the Transportation Security Administration (TSA) has been conducting random baggage checks on passengers at the country’s airports. The American Civil Liberties Union (ACLU) has recently begun to suspect that TSA staff at Boston’s Logan Airport are discriminatory in their selection of passengers for screenings. Assume that TSA Guidelines state that 1 in 18 passengers entering the departures gate should be selected for baggage screening on a random basis. The following table describes the actual number of passengers selected for screening on a randomly-selected day, broken down by various racial categories (“Other” includes “Mixed Race” and “Unknown”):

<table>
<thead>
<tr>
<th>Race</th>
<th>White</th>
<th>Black</th>
<th>Asian</th>
<th>Native Amer.</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td># Screened</td>
<td>99</td>
<td>20</td>
<td>9</td>
<td>0</td>
<td>13</td>
</tr>
</tbody>
</table>

Assume also that Logan Airport has reliable data on the overall racial breakdown of passengers at this time of year, and that the percentages are known to be the following:

<table>
<thead>
<tr>
<th>Race</th>
<th>Percentage of All Passengers</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>74</td>
</tr>
<tr>
<td>Black</td>
<td>9</td>
</tr>
<tr>
<td>Asian</td>
<td>5</td>
</tr>
<tr>
<td>Native American</td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td>11</td>
</tr>
</tbody>
</table>

Based on this information, is it reasonable to conclude that there is an association between a passenger’s race and whether or not he or she is selected for baggage screening by the TSA? Why or why not? Please conduct whatever statistical test(s) you think would most appropriate and interpret the results. Also, be sure to note and concerns or caveats you might have about your analysis.
5. You are a project manager for a City working on a downtown revitalization project which includes the reconstruction of a \(\frac{1}{2}\)-mile stretch of Main Street. The construction has been making steady progress, but the Mayor is eager for project to be completed, as the City’s Chief Financial Officer has just calculated that the project is costing the City $12,000 per day in non-construction expenses (police details, re-routing of traffic, compensation to area businesses for lost sales, and other mitigation costs).

Unfortunately, at a recent construction oversight meeting you learn of a potential problem which could delay the entire project: due to unforeseen soil conditions, the project engineer now recommends treating the concrete with a chemical additive before pouring, but the crew does not have any of this chemical at the job site. When pushed on this point, you learn the following:

- If you decide to wait, there is a chance (estimated at 30%) that the engineer will be able to obtain enough of the additive from a local supplier to allow you to proceed with the project, but it would require one week to explore this option; if no local sources can be found, the engineer is 100% certain (at least for the purposes of this exam) that in exactly five weeks from today his normal supplier will have it back in stock anyways and the project can continue.

- If you insist on pouring concrete without the additive the engineer estimates that there is a three-out-of-four chance that it will fail to set properly, resulting in an additional cost of $43,000 and a further four-week delay to re-prepare the site and re-pour the concrete (you can assume that the re-pour would set properly). (Note that this does of course mean there would be a one-in-four chance you could continue without additional delay, if you are willing to accept this risk.)

(Continued on next page.)
(a) Given all this information, construct a decision tree with expected values and recommend a course of action to the Mayor to minimize cost to the City. (For the moment, assume that cost figures are reliable, all estimates can be taken as actual probabilities, a week equals seven days, and that you have all the information that will ever be available to you to advise the Mayor.)

(b) Now, disregarding the comment in parenthesis in part (a) above: if this were a real situation, what additional information might you need to know?
6. The Occupational Safety and Health Administration (OHSA) has just completed the largest-ever study of the noise impacts of heavy machinery on workers. Among the findings, the study examined the effectiveness of the new “Zykon”-brand noise-reduction devices, which can be installed on new or existing heavy forklifts. The table below presents data on the number of cases of hearing damage experienced by workers operating or working in close proximity to forklifts fitted with this technology versus those without “Zykon”-brand devices:

<table>
<thead>
<tr>
<th>Group</th>
<th>Hearing damage</th>
<th>No damage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Zykon</td>
<td>679</td>
<td>2,076</td>
</tr>
<tr>
<td>Zykon</td>
<td>661</td>
<td>2,250</td>
</tr>
</tbody>
</table>

(a) Which group shows a lower percentage of hearing damage among workers?

(b) Is this difference significant? (In determining significance, please state your research hypothesis and the associated null-hypothesis you are evaluating.)

(c) When the results are published, a construction industry trade group argues that the Zykon-brand noise reduction devices are extremely expensive to install, citing an average cost of $26,000 per forklift. How would you respond to this concern, using the data you have available? What other questions might you ask.
7. A researcher investigating crime data in Florida has noticed an interesting correlation: counties with higher percentages of registered Democrats also have more violent crimes overall. The scatter plot below shows this data for all 67 counties, with the percentage of registered Democrats along the x-axis,¹ and the total number of violent crimes reported in the county in 2007 along the y-axis.²

![Scatter plot showing correlation between percentage of registered Democrats and total violent crimes in 2007 across Florida counties.]

The researcher runs a regression on this data and gets the following results (the regression line is also shown on the plot):

|                      | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------------|----------|------------|---------|----------|
| (Intercept)          | 1116.2   | 115.1      | 9.697   | 2.99e-14 |
| Dem. Percent         | 703.3    | 210.6      | 3.339   | 0.00139  |

Adjusted R-squared: 0.1333

¹Technically this is mislabelled—it is actually showing the proportion of registered Democrats; note that the variable runs from about .25 (i.e., 25%) to .85 (i.e., 85%).
²I just made up this data.

Name:
(a) In plain English, interpret the results of this regression. Be sure to explain the meanings of the intercept, slope, $r^2$, and p-values.

(b) What questions or concerns might you have about this data or this study? How might you modify the analysis to address these issues?
8. For the following graph, answer the following questions:

- What does the graph seem to be showing?
- What additional questions does the graph leave you with?
- How could this graph be made better?

[CNN News Channel Poll Graph]

Question 2: Based on what you have heard or read about the case, do you agree with the court’s decision to have the feeding tube removed?

Results by party

<table>
<thead>
<tr>
<th>Party</th>
<th>Agree</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democrats</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>Republicans</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>Independents</td>
<td>54</td>
<td></td>
</tr>
</tbody>
</table>

Sample: Interviews conducted by telephone March 18-20, 2005, with 900 adults in the United States. Sampling error: ±4.7% pts.