Test 2 Information

Logistics

- The exam will be given in SB 120 & 128.
- You will have access to the computers so that you can use R, sage, and Mathematica.
- Be sure to indicate how you use sage or your calculator on the exam. No mystery numbers or expressions, please.
- Use good notation and show all work. Full credit will not be given if your answers are not sufficiently supported.

Material Covered

Test 2 covers chapters 1–3 of the statistics text. It may also include a review question from line integrals (Sections 13.1-13.4) For more details about what we covered when, consult the web calendar.

Format issues

1. Any questions similar to those from the text book are, of course, reasonable for the test as well.
2. You should also be able to answer questions based on R output (numerical or graphical).
3. I will design the test in such a way that you do not have to print any R output. For example, you might use R to calculate something and then copy the output (numerical values, equation of a regression line, etc.) from the screen to your test paper or I might ask you to make a crude hand sketch of a plot that you make in R.

A few more details

1. Three kinds of distributions: discrete, continuous, data
2. Numerical summaries: mean, variance, standard deviation, quantiles (including median), IQR, trimmed mean for all three types of distribution.
3. Graphical summaries: stemplot, histogram, boxplot, quantile-quantile plot, scatter plot (including how to get R to make them)
4. Proportions (probabilities)
   (a) You should be able to work from a mass function or density function (using sums or integrals).
   (b) For familiar distributions (see table), you should know how to use the relevant functions in R (dnorm(), pnorm(), qnorm(), etc.).
5. Linear Models (Regression)
   (a) Correlation coefficient
   (b) Least squares best fit
   (c) Using lm() to fit a model in R
   (d) Residuals and residual plots
   (e) Transformations to improve fit
      - Transformations may be based on a priori information
      - Transformations may be motivated by the data itself (see Figure 3.15)
   (f) Using multiple predictors

©2009 Randall Pruim (rpruim@calvin.edu)
## Distributions Cheat Sheet

The following table will be printed on your exam.

<table>
<thead>
<tr>
<th>distribution</th>
<th>mass or density function</th>
<th>mean</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson</td>
<td>( dpois(x, \lambda) = e^{-\lambda} \frac{\lambda^x}{x!} )</td>
<td>( \lambda )</td>
<td>( \lambda )</td>
</tr>
<tr>
<td>Binomial</td>
<td>( dbinom(x,n,\pi) = \binom{n}{x}\pi^x(1-\pi)^{n-x} )</td>
<td>( n\pi )</td>
<td>( n\pi(1-\pi) )</td>
</tr>
<tr>
<td>Geometric</td>
<td>( dgeom(x,\pi) = \pi \frac{1}{1-\pi} )</td>
<td>( \frac{1}{\pi} - 1 )</td>
<td>( \frac{1-\pi}{\pi^2} )</td>
</tr>
<tr>
<td>Neg. Binomial</td>
<td>( dnbinom(x,size=s,prob=\pi) = \binom{x+n-1}{s}\pi^s(1-\pi)^x )</td>
<td>( \frac{s}{\pi} - s )</td>
<td>( \frac{s(1-\pi)}{\pi^2} )</td>
</tr>
</tbody>
</table>

| Uniform          | \( dunif(x,a,b) = \frac{1}{b-a} \) on \([a,b]\) | \( \frac{b+a}{2} \) | \( \frac{(b-a)^2}{12} \) |
| Std. normal      | \( dnorm(x,0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \) | 0 | 1                |
| Normal           | \( dnorm(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \) | \( \mu \) | \( \sigma^2 \) |
| Lognormal        | \( dlnorm(x,meanlog=\mu, sdlog=\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\ln(x)-\mu)^2/2\sigma^2} e^{\mu+\sigma^2/2} e^{2\mu+\sigma^2}(e^{\sigma^2}-1) \) |                |                  |
| Exponential      | \( dexp(x,\lambda) = \lambda e^{-\lambda x} \) | \( 1/\lambda \) | \( 1/\lambda^2 \) |
| Gamma            | \( dgamma(x,\alpha, rate=\lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \) | \( \alpha/\lambda \) | \( \alpha/\lambda^2 \) |
| Weibull          | \( dweibull(x,shape=\alpha, scale=\beta) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-x/\beta^\alpha} \) | \( \beta(1+\frac{1}{\alpha}) \) | \( \beta^2 \left[ \Gamma(1+\frac{1}{\alpha}) - \left[ \Gamma(1+\frac{1}{\alpha}) \right]^2 \right] \) |
| Beta             | \( dbeta(x,shape1=\alpha, shape2=\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \) | \( \frac{\alpha}{\alpha+\beta} \) | \( \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \) |

### Notes (These will not appear on the exam)

1. \( \Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \, dt \)
   \( \Gamma(x) \) can be evaluated using integration by parts, but you won’t ever need to do that. This is taken care of for you in functions like \( \text{dbeta}() \), \( \text{pbeta}() \), etc. If you ever need just the value of \( \Gamma(x) \) (for example, to get the mean of a Weibull distribution), you can use the function \( \text{gamma}() \) in R.

2. Exponential distributions are a special case of the Weibull distributions.

3. Exponential distributions are also a special case of the gamma distributions.

4. Geometric distributions are a special case of the negative binomial distributions.

©2009 Randall Pruim (rpruim@calvin.edu)