Learning Expectations for Final Exam

At the completion of the course, students should have learned the concepts listed below and be able to perform the specified tasks. The exam is cumulative, but emphasis will be placed on topics covered since Exam II (see **). See * for other topics to focus on.

Probability Theory

*P1. Students should know the 3 basic axioms for probability and the definitions and role of an experiment, the sample space, an event, and a sample point.

\[
\begin{align*}
P(A) &\geq 0; \\
P(S) &= 1; \\
\text{for } A \text{ and } B \text{ disjoint, } P(A \cup B) &= P(A) + P(B) \\
\end{align*}
\]

*P2. Students should know how to calculate the probability of events consisting of unions, intersections, and complements of events of known probability.

*P3. Students should know the formula for conditional probability and should be able to calculate the conditional probabilities of various events. Students should know the definition and use of independence. For independent events A and B, \( P(A \cap B) = P(A) \cdot P(B) \).

P4. Students should be able to recognize Total Probability & Bayes Theorem problems, structure the problem using a tree, and calculate \( P(E_i|A) \) and \( P(A) \), given \( P(E_i) \) and \( P(A|E_i) \) for all \( i \).

Counting

C1. Students should know when counting is an appropriate basis for determining probabilities.

<< All sample points equally likely. >>

C2. Students should know the formulas for

- ordered sets drawn with replacement \( [N = n^k] \)
- ordered sets without replacement \( [\text{permutations } N = n!/(n-k)!] \)
- unordered sets without replacement \( [\text{combinations } N = n!/(k!(n-k)!)] \)

C3. Students should be able to use the formulas from C2 and P2 to calculate the probability of various events for experiments related to ordered sets, job assignments, and committees.

Implementation of Probability Theory and Random Variables

*I1. Students should know the definition of a random variable, why the concept is useful, and how this concept is used in engineering applications.
*I2. Students should know definitions for univariate cdf, pmf, and pdf; the properties of each; their relationship; and how to use these functions to calculate percentiles and probabilities for events such as \( a \leq X \leq b \).

*I3. Students should know how to compute the mean and variance of a random variable for a given pdf or pmf, and how to compute such moments for linear functions and linear combinations of random variables.

Distributions

D1. Students should recognize when the characteristics of a discrete counting problem leads to a binomial distribution (counting with finite \( n \) and given \( p \)), and be able to use the information provided in a problem and knowledge of the properties of the distribution to calculate the mean and variance of a random variable, as well as the probabilities of various values and events.

*D2. Students should know the general properties of a normal distribution (symmetric, additive \( a + bX \) or \( aX + bY \), unbounded, cdf unavailable in closed form), the form of the pdf, and how to calculate percentiles and the probability of events such as \( a \leq X \leq b \) for \( X \sim N[\mu, \sigma^2] \).

D3. Students should be able to calculate means, variances, percentiles and other probabilities associated with the uniform and exponential distributions.

*D4. Students should be able to state the Central Limit Theorem and know when it applies.

*D5. Students should be able to calculate the moments and percentiles of a lognormal distribution from its parameters (the log-space mean and variance). They should be able to compute the parameters given the mean and variance of the distribution, or some mixture of percentiles and moments. This requires knowing what equations relate the various parameters, moments, and percentiles; and what various symbols represent.

D6. Students should be able to calculate the parameters of the Gumbel distribution given the mean and variance. They should also be able to calculate means, variances, percentiles and other probabilities associated with the distribution for specified parameter values.

D7. Students should be able to calculate (at least one of) the parameters of the Weibull distribution given the mean and variance. They should also be able to calculate means, variances, percentiles and other probabilities associated with the distribution for specified parameter values.
CE 3710: Uncertainty Analysis in Engineering

Estimators
*E1. Students should know why the concept of an estimator is important, and appreciate how estimators become more accurate with larger sample sizes.

*E2. Students should understand the sampling properties of estimators of the mean and variance of a random variable.

*E3. Students should be able to derive method of moment estimators for simple cases, thus showing they understand how the basic principles apply.

Confidence Intervals
*L1. Students should understand what a confidence interval is and how they quantify uncertainty.

*L2. Students should know how to construct confidence intervals for a set of standard cases: mean of normal distribution with small samples (σ known and unknown), mean with large samples, and the standard deviation for the normal distribution.

Hypothesis Testing
*H1. Students should know how to structure a statistical decision problem as a choice between two hypotheses and how that choice relates to type I and II errors (α and β). They should also be able to articulate what these errors represent.

*H2. Students should know how to select the null and alternative hypotheses to achieve the intended purpose of a test, and thus establish upon which hypothesis (Hₐ) the burden of proof should fall, and which will be accepted if little data is available (H₀).

**H3. Students should be able to calculate the test statistic, rejection region for a specified α, and make a decision for one- and two-sample z- and t-tests. Students should also recognize when a paired analysis is appropriate and be able to perform the associated hypothesis test.

Goodness-of-Fit Analysis
G1. Students should know how to assign plotting positions to a data set and to construct an empirical CDF

*G2. Students should be able to construct both probability plots and quantile-quantile plots, and they should be understand how such a plot illustrates whether the data is drawn from the postulated distribution (e.g., normal, lognormal, Gumbel).

**G3. Students should be able to compute the probability plot correlation test statistic, and perform a hypothesis test (Ryan Joiner Test of Normality) of whether the data set (or some transformation thereof) was drawn from a normal distribution.
Regression

R1. Students should understand the form of and assumptions employed with the simple linear model $Y = \beta_0 + \beta_1 x + \epsilon$, with independent and normal errors $\epsilon$.

R2. Students should be able to calculate least-squares estimators of $\beta_0$ and $\beta_1$, estimate the precision of these estimators, and conduct relevant hypothesis tests.

R3. Students should be able to calculate and explain the meaning of confidence intervals for estimators of $\beta_0$ and $\beta_1$, and the mean of $y$ for a given $x$, and of prediction intervals for a future observation of $y$.

R4. For the simple linear model, students should know what $R^2$ and adjusted-$R^2$ represent and how they are calculated. They should also know the meaning of a sum-of-squares (ANOVA) table and the residual sum of squares.

R5. Students should know the definition and meaning of the correlation coefficient, be able to calculate its estimator $R_{xy}$, and be able to conduct a hypothesis test of whether $\rho = 0$.

R7. Students should also recognize how many nonlinear models can be transformed into models which can be analyzed as linear models.

For example: $y = ax^b$ becomes $\ln y = \ln a + b \ln x$

Sample Exam Questions

1. A sample of 12 observations on $X$ and $Y$ yields a correlation coefficient of 0.682. Using a level of significance of 5%, test the null hypothesis that the correlation coefficient of the population is 0.0.

   Answer: Test $H_0: \rho_{xy} = 0$ vs. $H_a: \rho_{xy} \neq 0$; (b) $T = 2.949$ with $v = 10$; Reject $H_0$ if $T \geq 2.228$ or $T \leq -2.228$; Reject $H_0$

2. A platform is supported by 10 columns. For each column, the load bearing capacity is normally distributed with mean 50 tons and standard deviation 4.5 tons. Assume the load bearing capacities of the columns are independent of one another and the platform’s capacity is the sum of the capacities of the 10 columns. What design load can an engineer use and be 99% sure the platform will support that value?

   Answer: $W = \sum X_i \sim N[500, 14.2^2]$ where $X_i \sim N[50, 4.5^2]$; $W_{0.01} = 467$
3. Students have been casting concrete cylinders to test the strength of a particular mix. Here are the results from 12 samples:

5610, 4440, 3430, 5760, 6320, 5640, 2025, 4275, 4005, 6530, 5240, 4120

To evaluate if the strengths are normally distributed, the students computed the correlation between the ranked observations and the corresponding standard normal percentiles. They obtained $R = 0.974$. For $\alpha = 10\%$, conduct a probability plot correlation test to see if the normal distribution is a reasonable model for the data.

*Answer: $0.9347 < \rho_c < 0.9506$, therefore, fail to reject $H_0$ – the normal distribution is appropriate*

4. In watershed studies, researchers often develop a physical watershed model using software such as HEC-HMS or SWAT. The model must be calibrated to observed flows, typically on a daily time step over an extended duration. The objective during the calibration process is to minimize the differences between the observed flows and those estimated by the model. Consider a new model that was developed using data for 15 storms that had at least 2 inches of rainfall. The observed and estimated peak flow rates are reported below. While the model does not exactly reproduce the peak for each storm for a number of reasons, including simplifications in the model and the precision of the data, a concern is whether the model tends to either overestimate or underestimate peak flows.

<table>
<thead>
<tr>
<th>Storm Event</th>
<th>Observed Peak Flow (x)</th>
<th>Estimated Peak Flow (y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>113.0</td>
<td>90.6</td>
</tr>
<tr>
<td>2</td>
<td>30.2</td>
<td>34.0</td>
</tr>
<tr>
<td>3</td>
<td>62.8</td>
<td>68.5</td>
</tr>
<tr>
<td>4</td>
<td>351.8</td>
<td>283.9</td>
</tr>
<tr>
<td>5</td>
<td>45.9</td>
<td>35.1</td>
</tr>
<tr>
<td>6</td>
<td>140.4</td>
<td>105.7</td>
</tr>
<tr>
<td>7</td>
<td>97.2</td>
<td>75.5</td>
</tr>
<tr>
<td>8</td>
<td>18.4</td>
<td>17.6</td>
</tr>
<tr>
<td>9</td>
<td>54.7</td>
<td>38.3</td>
</tr>
<tr>
<td>10</td>
<td>71.1</td>
<td>49.3</td>
</tr>
<tr>
<td>11</td>
<td>42.5</td>
<td>29.4</td>
</tr>
<tr>
<td>12</td>
<td>104.5</td>
<td>92.1</td>
</tr>
<tr>
<td>13</td>
<td>111.1</td>
<td>118.2</td>
</tr>
<tr>
<td>14</td>
<td>52.1</td>
<td>29.5</td>
</tr>
<tr>
<td>15</td>
<td>270.8</td>
<td>258.6</td>
</tr>
<tr>
<td><strong>Average:</strong></td>
<td><strong>104.44</strong></td>
<td><strong>88.44</strong></td>
</tr>
<tr>
<td><strong>Standard Deviation:</strong></td>
<td><strong>92.03</strong></td>
<td><strong>80.52</strong></td>
</tr>
</tbody>
</table>

(a) What are the appropriate hypotheses to determine if a significant difference exists in the mean peak flow rate estimated by the model versus that observed?
CE 3710: Uncertainty Analysis in Engineering

(b) What is the value of the test statistic for this data set?

(c) What is the rejection region for a test with a type I error of 5%?

(d) What do you conclude for this data?

For the set of observed peak flow rates provided above:

(e) Construct a 95% confidence interval for the true mean peak flow rate.

(f) Construct a 95% confidence interval for the true standard deviation of the peak flow rate.

\textit{Answer:} (a) \( H_0: \mu_D = 0, \ H_1: \mu_D \neq 0; \) (b) \( T = 3.320 \) with \( v = 14; \)

(c) Reject \( H_0 \) if \( T \geq 2.145 \) or if \( T \leq -2.145; \) (d) Reject \( H_0; \) (e) (53.47, 155.41); (f) (67.38, 145.14)

\textit{NOTE: This data really should be paired. If you incorrectly considered the samples to be independent, then the answers to the above questions would be as follows:}

(a) \( H_0: \mu_x = \mu_y, \ H_1: \mu_x \neq \mu_y; \) (b) \( T = 0.507 \) with \( v = 27; \) (c) Reject \( H_0 \) if \( T \geq 2.052 \) or if \( T \leq -2.052; \)

(d) Fail to reject \( H_0 \)

5. A geotechnical engineer is interested in the shear strength of soil in a clay stratum at various depths. She obtained 9 measurements of shear strengths (in kips per square foot – ksf) at 9 depths:

\begin{align*}
\text{Depth } x: & \quad \bar{x} = 6.044 \text{ m} \quad sx = 2.7220 \text{ m} \\
\text{Strength } y: & \quad \bar{y} = 0.911 \text{ ksf} \quad sy = 0.428 \text{ ksf} \\
\sum(x_i - \bar{x})(y_i - \bar{y}) = 8.196
\end{align*}

(a) What are the least-squares estimates of \( \beta_0 \) and \( \beta_1 \) for the simple linear model

\[ Y = \beta_0 + \beta_1 x + \varepsilon \]

relating depth \( x \) with shear strength \( Y \), with errors \( \varepsilon \) due to soil variability and laboratory measurement precision?

(b) Estimate of the residual mean square error \( s^2_e \).

(c) Estimate the variances of the least-squares estimators of \( \beta_0 \) and \( \beta_1 \).

(d) What are the values of \( R^2 \) and the adjusted-\( R^2 \) for this data set?

(e) Before gathering this data the geotechnical engineer believed that shear strength should increase with depth due to the structure and clay content of the stratum. With this presupposition, what are the appropriate null and alternative hypotheses on \( \beta_0 \)? What is the rejection region for the 5% test of whether \( \beta_1 = 0 \)? Is the engineer’s prior belief supported by this data?
(f) Using your fitted line, what is the best estimate of the value of y for a depth of x = 7.5 meters? What is a 95% prediction interval for the shear strength that would be obtained for a future measurement made at that depth?

Answer: (a) \( b_0 = 0.077, b_1 = 0.138 \); (b) 0.0478; (c) \( \text{Var}(b_0) = 0.0348, \text{Var}(b_1) = 0.000806 \); (d) 0.772, 0.739; (e) \( H_0: \beta_1 = 0, H_1: \beta_1 > 0 \); Reject \( H_0 \) if \( T \geq 1.895 \); \( T = 4.861 \), reject \( H_0 \)--yes; (f) 1.112 ksf, (0.558, 1.666)

6. A geotechnical engineer is studying the effect of soil properties on the earthquake damage experienced by civil infrastructure including pipelines and buildings. Let \( E \) be a measure of likely earthquake intensity that reflects soil properties, and \( D \) a measure of the damage to residential houses in a quake. Here are some summary statistics based on data the engineer has collected for \( n = 24 \) events:

<table>
<thead>
<tr>
<th>Earthquake Intensity</th>
<th>E: ( \bar{E} = 67.71 )</th>
<th>( s_E = 57.50 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Damage D:</td>
<td>( \bar{D} = 17.72 )</td>
<td>( s_D = 14.28 )</td>
</tr>
<tr>
<td>( \sum (E_i - \bar{E})(D_i - \bar{D}) )</td>
<td>11,400</td>
<td></td>
</tr>
</tbody>
</table>

(a) What are the least-squares estimates of \( \beta_0 \) and \( \beta_1 \) for the simple linear model

\[ D = \beta_0 + \beta_1 E + \varepsilon \]

where \( \varepsilon \) is the model error?

(b) Estimate of the residual mean square error \( s^2_{\varepsilon} \).

(c) What portion of observed variability in \( D \) is explained by \( E \)?

(d) Estimate the variance of the least-squares estimator of \( \beta_1 \).

(e) Before gathering this data the geotechnical engineer believed that damage should increase with earthquake intensity. With this presupposition, what are the appropriate null and alternative hypotheses on \( \beta_1 \)? What is the rejection region for the 5% test of whether \( \beta_1 = 0 \)? Is the engineer’s prior belief supported by this data?

(f) Using your fitted line, what is a 90% prediction interval for the damage at a site with an earthquake intensity \( E = 120 \)?

Answer: (a) \( b_0 = 7.564, b_1 = 0.150 \); (b) 135.5; (c) 0.364; (d) (0.042)²; (e) \( H_0: \beta_1 = 0, H_1: \beta_1 > 0 \); Reject \( H_0 \) if \( T \geq 1.717 \); \( T = 3.552 \), reject \( H_0 \)--yes; (f) (4.82, 46.32)

Page 7