ESD.86
Exam #2 Review

Dan Frey
Associate Professor of Mechanical Engineering and Engineering Systems
Some Study Suggestions

• Review all the lecture notes
  – If there is a concept test, know the answer and WHY it's the right answer and WHY other answers are wrong
  – If there is a word you don't understand, look it up, talk to colleagues...

• Review the last two homeworks
  – For concepts, not details

• Review the reading assignments (April and May)
  – For concepts, not details
  – If there is a graph or table, make sure you can describe how to use it and the procedure by which it was made
Suggested Resources
(I would not emphasize this as much as the lecture notes and problem sets)

Wu and Hamada. *Experiments: Planning, Analysis and Parameter Design Optimization*. Chapters 4 and 5 are relevant. The "practical summaries" are good condensations of the main points. Many of the exercises are close to what you might find on the test. Many are not e.g. "prove that ..." or "find the defining words and resolution..."

*Problem Solvers: Statistics* (Research & Education Association)
Solving lots of problems is good practice. There are books filled with problems and solutions. A large fraction of these involve lots of number crunching, these are fine to review but that's not the sort of question I intend to give. Many are conceptual or involve modest computations or estimation. Those are closer in style to what you can expect. There are many topics we didn't cover so you don't have to study them.
Weibull’s Derivation

Call $P_n$ the probability that a chain will fail under a load of $x$

Let’s define a cdf for each link meaning the link will fail at a load $X$ less than or equal to $x$ as $P(X \leq x) = F(x)$

If the chain does not fail, it’s because all $n$ links did not fail

If the $n$ link strengths are probabilistically independent

$$1 - P_n = (1 - P)^n$$

Some Terms Related to Estimation

- **Consistent** – for any \( c \), \( \lim_{n \to \infty} P \left( \left| \hat{\theta} - \theta \right| \geq c \right) = 0 \)  
- **Unbiased** – \( E(\hat{\theta}) = \theta \)  
- **Minimum variance**

\[
\text{var}(\hat{\theta}) = \frac{1}{nE} \left[ \left( \frac{\partial \ln f(X)}{\partial \theta} \right)^2 \right]
\]

**MLEs** are not always close. **MLEs** are not always close.
Complex Distributions

Looking for Further Evidence of Two Populations

No evidence of bimodality in fatigue data

Clear evidence of bimodality in strength data
Reliability Terminology

• Reliability function \( R(t) \) -- The probability that a product will continue to meet its specifications over a time interval

• Mean Time to Failure \( MTTF \) -- The average time \( T \) before a unit fails

\[
MTTF = \int_0^\infty R(t)dt
\]

• Instantaneous failure rate \( \lambda(t) \)

\[
\lambda(t) = \Pr(\text{System survives to } t + dt \mid \text{System survives to } t)
\]

\[
R(t) = e^{-\int_0^t \lambda(\xi)d\xi}
\]
“When the system operating time is the MTBF, the reliability is 37%”
- Blanchard and Fabrycky

\[ R(t) = e^{-\lambda t} \]
Fisher's Null Hypothesis Testing

1. Set up a statistical null hypothesis. The null need not be a nil hypothesis (i.e., zero difference).
2. Report the exact level of significance ... Do not use a conventional 5% level, and do not talk about accepting or rejecting hypotheses.
3. Use this procedure only if you know very little about the problem at hand.
Gigernezer's Quiz

Suppose you have a treatment that you suspect may alter performance on a certain task. You compare the means of your control and experimental groups (say 20 subjects in each sample). Further, suppose you use a simple independent means $t$-test and your result is significant ($t = 2.7$, $d.f. = 18$, $p = 0.01$). Please mark each of the statements below as “true” or “false.” ...

1. You have absolutely disproved the null hypothesis
2. You have found the probability of the null hypothesis being true.
3. You have absolutely proved your experimental hypothesis (that there is a difference between the population means).
4. You can deduce the probability of the experimental hypothesis being true.
5. You know, if you decide to reject the null hypothesis, the probability that you are making the wrong decision.
6. You have a reliable experimental finding in the sense that if, hypothetically, the experiment were repeated a great number of times, you would obtain a significant result on 99% of occasions.
Concept Question

• This Matlab code repeatedly generates and tests simulated "data"
• 20 "subjects" in the control and treatment groups
• Both normally distributed with the same mean
• How often will the $t$-test reject $H_0$ ($\alpha=0.01$)?

for i=1:1000
    control=random('Normal',0,1,1,20);
    trt=random('Normal',0,1,1,20);
    reject_null(i) = ttest2(control,trt,0.01);
end
mean(reject_null)

1) ~99% of the time
2) ~1% of the time
3) ~50% of the time
4) None of the above
Concept Question

- This Matlab code repeatedly generates and tests simulated "data"
- 20 "subjects" in the control and treatment groups
- Both normally distributed with the different means
- How often will the \( t \)-test reject \( H_0 (\alpha=0.01) \)?

```matlab
for i=1:1000
    control=random('Normal',0,1,1,200);
    trt=random('Normal',1,1,1,200);
    reject_null(i) = ttest2(control,trt,0.01);
end
mean(reject_null)
```

1) ~99% of the time
2) ~1% of the time
3) ~50% of the time
4) None of the above
Concept Question

• How do “effect” and "alpha" affect the rate at which the \( t \)-test rejects \( H_0 \)?

```matlab
effect=1;alpha=0.01;
for i=1:1000
    control=random('Normal',0,1,1,20);
    trt=random('Normal',effect,1,1,20);
    reject_null(i) = ttest2(control,trt,alpha);
end
mean(reject_null)
```

a) ↑ effect, ↑ rejects
b) ↑ effect, ↓ rejects
c) ↑ alpha, ↑ rejects
d) ↑ alpha, ↓ rejects

1) a & c
2) a & d
3) b & c
4) b & d
NP Framework and Two Types of Error

- Set a critical value $c$ of a test statistic $T$ or else set the desired confidence level or "size" $\alpha$
- Observe data $X$
- Reject $H_1$ if the test statistic $T(X) \geq c$
- Probability of Type I Error – The probability of $T(X) < c \mid H_1$
  – (i.e. the probability of rejecting $H_1$ given $H_1$ is true)
- Probability of Type II Error – The probability of $T(X) \geq c \mid H_2$
  – (i.e. the probability of not rejecting $H_1$ given $H_2$ is true)
- The power of a test is $1 - \text{probability of Type II Error}$
- In the N-P framework, power is maximized subject to Type I error being set to a fixed critical value $c$ or of $\alpha$
Measures of Central Tendency

- **Arithmetic mean**
  - an unbiased estimate of
  \[ \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \]

- **Median**
  \[ \mu = E(x) = \int_{S} xf_x(x)dx \]
  
  \[= \left\{ \begin{array}{ll}
    X_{n+1} & \text{if } n \text{ is odd} \\
    \frac{n+1}{2} & \\
    \frac{1}{2} \left( X_{n/2} + X_{1+n/2} \right) & \text{if } n \text{ is even}
  \end{array} \right. \]

- **Mode** – The most frequently observed value in the sample
Confidence Intervals

- Assuming a given distribution and a sample size $n$ and a given value of the parameter $\theta$, the 95% confidence interval from $U$ to $V$ is s.t. the estimate of the parameter $\hat{\theta}$

  $$\Pr(U < \hat{\theta} < V | \theta) = 95\%$$

- The confidence interval depends on the confidence level, the sample variance, and the sample size.
Measures of Dispersion

- Population Variance
  \[ VAR(X) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \]

- Sample variance
  \[ S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \]
  – an unbiased estimate of \[ \sigma^2 = E((x - E(x))^2) \]

- \( n^{th} \) central moment
  \[ E((x - E(x))^n) \]

- \( n^{th} \) moment about \( m \)
  \[ E((x - m)^n) \]
Skewness and Kurtosis

- **Skewness** \[ E((x - E(x))^3) \]

  Positively skewed distribution

- **Kurtosis** \[ E((x - E(x))^4) \]

  Positive kurtosis
Correlation Coefficient

• Sample

\[ r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)S_X S_Y} \]

\[ S_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \]

• Which is an estimate of

\[ \frac{E((x - E(x))(y - E(y))))}{\sigma_x \sigma_y} \]
But What Does it Mean?

Pick a Box! Are they all the same?

Correlation 0.7
Correlation 0.7
Correlation 0.7
Correlation 0.7

What is Linear Regression?

1. Form a probabilistic model

\[ Y = \alpha + \beta x + \varepsilon \]

2. Get a sample of data in pairs \((X_i, Y_i), i=1\ldots n\)
3. Estimate the parameters of the model from the data
The Method of Least Squares

Given a set of $n$ data points $(x, y)$ pairs

There exists a unique line $\hat{y} = a + bx$ that minimizes the residual sum of squares

$$\sum_{i=1}^{n} e_i^2$$

$$e_i = y_i - \hat{y}_i$$

$$s_e^2 = \frac{1}{n-2} \sum_{i=1}^{n} e_i^2$$
Matlab Code for Regression

\[ p = \text{polyfit}(x, Y, 1) \]
\[ y\_hat = \text{polyval}(p, x); \]
\[ \text{plot}(x, y\_hat, '-', 'Color', 'g') \]
Concept Question

You are seeking to calibrate a load cell. You wish to determine the regression line relating voltage (in Volts) to force (in Newtons). What are the units of $a$, $b$, $S_{xx}$, and $S_{xy}$ respectively?

1) N, N, N, and N
2) V, V, V$^2$, and V$^2$
3) V, V/N, N$^2$, and VN
4) V/N, N, VN, and V$^2$
5) None of the variables have units
Regression Curve vs Prediction Equation

\[ Y = \alpha + \beta x + \varepsilon \]

Regression Curve

\[ \hat{y} = a + bx \]

Prediction Equation

Random \( E(\varepsilon) = 0 \)

estimated \( E(Y|x) \)

computed estimates of theoretical parameters \( \alpha \) and \( \beta \)
Evaporation vs Air Velocity Hypothesis Tests

<table>
<thead>
<tr>
<th>Air vel (cm/sec)</th>
<th>Evap coeff. (mm²/sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.18</td>
</tr>
<tr>
<td>60</td>
<td>0.37</td>
</tr>
<tr>
<td>100</td>
<td>0.35</td>
</tr>
<tr>
<td>140</td>
<td>0.78</td>
</tr>
<tr>
<td>180</td>
<td>0.56</td>
</tr>
<tr>
<td>220</td>
<td>0.75</td>
</tr>
<tr>
<td>260</td>
<td>1.18</td>
</tr>
<tr>
<td>300</td>
<td>1.36</td>
</tr>
<tr>
<td>340</td>
<td>1.17</td>
</tr>
<tr>
<td>380</td>
<td>1.65</td>
</tr>
</tbody>
</table>

Regression Statistics:
- Multiple R: 0.934165
- R Square: 0.872665
- Adjusted R Square: 0.854474
- Standard Error: 0.159551
- Observations: 9

ANOVA:
- df: 1, 7, 8
- SS: 1.221227, 0.178196, 1.399422
- MS: 1.221227, 0.025457, 0.159551
- F: 47.97306
- Significance F: 0.000226

Coefficients:
- Intercept: 0.102444 (P-value: 0.369673)
- X Variable 1: 0.003567 (P-value: 0.000226)

Residuals:
- Predicted Y:
  - Observation 1: 0.173778
  - Observation 2: 0.316444
  - Observation 3: 0.459111
  - Observation 4: 0.601778
  - Observation 5: 0.744444
  - Observation 6: 0.887111
  - Observation 7: 1.029778
  - Observation 8: 1.172444
  - Observation 9: 1.315111

- Residuals:
  - Observation 1: 0.006222
  - Observation 2: 0.053556
  - Observation 3: -0.10911
  - Observation 4: 0.178222
  - Observation 5: -0.18444
  - Observation 6: -0.13711
  - Observation 7: 0.150222
  - Observation 8: 0.187556
  - Observation 9: -0.14511

- Standard Residuals:
  - Observation 1: 0.041691
  - Observation 2: 0.35884
  - Observation 3: -0.73108
  - Observation 4: 1.19415
  - Observation 5: -1.23584
  - Observation 6: -0.91869
  - Observation 7: 1.19415
  - Observation 8: 1.25669
  - Observation 9: -0.97229

- Percentile Y:
  - Observation 1: 5.555556
  - Observation 2: 16.66667
  - Observation 3: 27.77778
  - Observation 4: 38.88889
  - Observation 5: 50
  - Observation 6: 61.11111
  - Observation 7: 72.22222
  - Observation 8: 83.33333
  - Observation 9: 94.44444

Normal Probability Plot:
- Sample Percentile Y:
  - Observation 1: Sample Percentile Y
  - Observation 2: Sample Percentile Y
  - Observation 3: Sample Percentile Y
  - Observation 4: Sample Percentile Y
  - Observation 5: Sample Percentile Y
  - Observation 6: Sample Percentile Y
  - Observation 7: Sample Percentile Y
  - Observation 8: Sample Percentile Y
  - Observation 9: Sample Percentile Y
Bayes' Theorem

\[ \Pr(A|B) \equiv \frac{\Pr(A \cap B)}{\Pr(B)} \]

\[ \Pr(B|A) \equiv \frac{\Pr(A \cap B)}{\Pr(A)} \]

with a bit of algebra

\[ \Pr(A|B) = \frac{\Pr(A) \Pr(B|A)}{\Pr(B)} \]
False Discovery Rates

Single Factor Experiments

- A single **experimental factor** is varied
- The parameter takes on various **levels**

<table>
<thead>
<tr>
<th>Cotton weight percentage</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>a=5 replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>7</td>
<td>7</td>
<td>15</td>
<td>11</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>12</td>
<td>17</td>
<td>12</td>
<td>18</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>14</td>
<td>18</td>
<td>18</td>
<td>19</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>19</td>
<td>25</td>
<td>22</td>
<td>19</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>7</td>
<td>10</td>
<td>11</td>
<td>15</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

Each cell is a $y_{ij}$

Each row is a treatment $i$

Fiber strength in lb/in$^2$
Breakdown of Sum Squares

"Grand Total Sum of Squares"

\[ GTSS = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij}^2 \]

"Total Sum of Squares"

\[ SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^2 \]

\[ SS_{Treatments} = n \sum_{i=1}^{a} (\bar{y}_{i.} - \bar{y}_{..})^2 \]

SS due to mean
\[ = n\bar{y}_{..}^2 \]

SS due to mean
\[ = n\bar{y}_{..}^2 \]
Breakdown of DOF

- \( N \): number of \( y \) values
- 1: due to the mean
- \( N-1 \): total sum of squares
- \( a-1 \): for the treatments
- \( N-a \): for error
What is a "Degree of Freedom"?

- How many scalar values are needed to unambiguously describe the state of this object?

- What if I were to fix the $x$ position of a corner?
What is a "Degree of Freedom"?

- How many scalar values are needed to unambiguously describe the outcome of this experiment?

- What if I were to tell you \( \bar{y} \)?
- What if I were to tell you \( \bar{y}_i \) \( i = 1 \ldots 4 \)?

<table>
<thead>
<tr>
<th>Cotton weight percentage</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>25</td>
<td>14</td>
</tr>
<tr>
<td>30</td>
<td>19</td>
</tr>
<tr>
<td>35</td>
<td>7</td>
</tr>
</tbody>
</table>
Adding h.o.t. to the Model Equation

Each row of $X$ is paired with an observation

There are $n$ observations of the response

You can add interactions

You can add curvature
Breakdown of Sum Squares

“Grand Total Sum of Squares”

\[ \text{GTSS} = \sum_{i=1}^{n} y_i^2 \]

\[ \text{SS due to mean} = n\bar{y}^2 \]

\[ \text{SS}_T = \sum_{i=1}^{n} (y_i - \bar{y})^2 \]

\[ \text{SS}_R = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \]

\[ \text{SS}_E = \sum_{i=1}^{n} e_i^2 \]

\[ \text{SS}_{PE} \]

\[ \text{SS}_{LOF} \]
Breakdown of DOF

- $n$: number of $y$ values
- $1$: due to the mean
- $n-1$: total sum of squares
- $k$: for the regression
- $n-k-1$: for error
Estimation of the Error Variance $\sigma^2$

Remember the model equation $y = X\beta + \varepsilon$

If assumptions of the model equation hold, then

$$E\left( SS_E / (n - k - 1) \right) = \sigma^2$$

So an unbiased estimate of $\sigma^2$ is

$$\hat{\sigma}^2 = SS_E / (n - k - 1)$$
Test for Significance
Individual Coefficients

The hypotheses are

\[ H_0 : \beta_j = 0 \]

\[ H_1 : \beta_j \neq 0 \]

The test statistic is

\[ t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \]

\[ C = \left( X^T X \right)^{-1} \]

Reject \( H_0 \) if \( |t_0| > t_{\alpha/2, n-k-1} \)
Factorial Experiments

Cuboidal Representation

Exhaustive search of the space of discrete 2-level factors is the full factorial $2^3$ experimental design.
Adding Center Points

Center points allow an experimenter to check for curvature and, if replicated, allow for an estimate of pure experimental error.
Concept Test

• You perform a linear regression of 100 data points \((n=100)\). There are two independent variables \(x_1\) and \(x_2\). The regression \(R^2\) is 0.72. Both \(\beta_1\) and \(\beta_2\) pass a \(t\) test for significance. You decide to add the interaction \(x_1x_2\) to the model. Select all the things that cannot happen:

1) Absolute value of \(\beta_1\) decreases
2) \(\beta_1\) changes sign
3) \(R^2\) decreases
4) \(\beta_1\) fails the \(t\) test for significance
Screening Design

- What is the objective of screening?
- What is special about this matrix of 1s and -1s?

Image removed due to copyright restrictions.

TABLE 2: Design I: Layout and Data for \(2^{8-4}\) Screening Design in Box and Liu, 1999.
Analysis of Variance

- What would you conclude about lack of fit?
- What is being used as the denominator of $F$?

Image removed due to copyright restrictions.

Concept Question

Say the independent experimental error of observations \((a), (ab), \) et cetera is \(\sigma_\varepsilon\).

We define the main effect estimate \(A\) to be

\[
A = \frac{1}{4}[(abc) + (ab) + (ac) + (a) - (b) - (c) - (bc) - (1)]
\]

What is the standard deviation of the main effect estimate \(A\)?

1) \(\sigma_A = \frac{1}{2}\sqrt{2}\sigma_\varepsilon\)  
2) \(\sigma_A = \frac{1}{4}\sigma_\varepsilon\)  
3) \(\sigma_A = \sqrt{8}\sigma_\varepsilon\)  
4) \(\sigma_A = \sigma_\varepsilon\)
Three Level Factors

8 vertices +
12 edges +
6 faces +
1 center =
27 points

$3^3$ Design
Factor Effect Plots

A

B

- 

+ 

30

52

20

40

B+

B-

A-

A+

A

B

0

- 

+ 

0

A
If there are no interactions in this system, then the factor effect plot from this design could look like:

Hold up all cards that apply.
A bracket holds a component as shown. The dimensions are strongly correlated random variables with standard deviations as noted. Approximately what is the standard deviation of the gap?

A) 0.011”  \( \sigma = 0.01" \)
B) 0.01”  \( \sigma = 0.001" \)
C) 0.009”
D) not enough info
Monte Carlo Simulations
What are They Good at?

\[
\text{Accuracy} \propto \frac{1}{\sqrt{N}} \quad N \equiv \# \text{Trials}
\]

• Above formulae apply regardless of dimension
• So, Monte Carlo is good for:
  – Rough approximations or
  – Simulations that run quickly
  – Even if the system has many random variables

Sampling Techniques for Computer Experiments

- **Random Sampling**
  - Clump
  - Gap

- **Stratified Sampling**

- **Latin Hypercube Sampling**
Errors in Scientific Software

• Experiment T1
  – Statically measured errors in code
  – Cases drawn from many industries
  – ~10 serious faults per 1000 lines of commercially available code

• Experiment T2
  – Several independent implementations of the same code on the same input data
  – One application studied in depth (seismic data processing)
  – Agreement of 1 or 2 significant figures on average

Next Steps

• Monday 7 May – Frey at NSF
• Wednesday 9 May – Exam #2
• Wed and Fri, May 14 and 16
  – Final project presentations