STAT 481 -- Final Practice Problems

1.Out of 25 women who had uterine cancer, 20 claimed to have used estrogens. Out of 30 women without uterine cancer 5 claimed to have used estrogens.

Exposure (estrogens)	Outcome (Cancer)	
	Yes	No
Yes	20	5
No	5	25

Test if there are significant association between the outcome and the exposure.

2. An article describes an experiment to determine the effect of C_2F_6 flow rate on the uniformity of the etch on a silicon wafer. Data for two flow rates are as follows:

C_2F_6	Uniformity Observation						N	Mean	StDev
	1	2	3	4	5	6			
125	2.7	4.6	2.6	3.0	3.2	3.8	6	3.317	0.760
200	4.6	3.4	2.9	3.5	4.1	5.1	6	3.933	0.821

a). Does the flow rate affect the wafer-to-wafer variability in etch uniformity? Use $\alpha = 0.05$.

b). Does the flow rate affect average uniformity? Use $\alpha = 0.05$.

3. The tensile strength of a paper product is related to the amount of hardwood in the pulp. Ten samples produced in the pilot plant, the data obtained are shown in the table:

Strength (x)	% Hardwood (y)	Strength (x)	% Hardwood (y)
160	10	181	20
171	15	188	25
175	15	193	25
182	20	195	28
184	20	200	30

$\Sigma x = 1829$
$\Sigma y = 208$

$\Sigma x^2 = 335825$
$\Sigma y^2 = 4684$
$\Sigma xy = 38715$
$n=10$

(a) Fit a linear regression model relating strength to percent hardwood.
(b) Test the model in part (a) for significance of regression.

4. The tensile strength of portland cement is being studied. Four different mixing techniques can be economically. The following data have been collected:

Mixing Technique	Tensile Strength (ib/in^{2}')			
1	3129	3000	2865	2890
2	3200	3300	2975	3150
3	2800	2900	2985	3050
4	2600	2700	2600	2765

Sample mean		Sum of Squares:
1	2971	
2	3156	SSTO=$6.436*10^5$
3	2933	SSTR=$4.897*10^5$
4	2666	

(a) Test the hypothesis that mixing techniques affect the strength of the cement. Use $\alpha = 0.05$.

(b). Compare the mean tensile strengths of the second and the fourth mixing techniques.

5. A textile mill has a large number of looms. Each loom is supposed to provide the same output of cloth per minute. To investigate the assumption, five looms are chosen at random and their output is noted at different time:

Loom	Output (lb/min)				
1	14.0	14.1	14.2	14.0	14.1
2	13.9	13.8	13.9	14.0	14-0
3	14.1	14.2	14.1	14.0	13.9
4	13.6	13.8	14.0	13.9	13.7
5	13.8	13.6	13.9	13.8	14.0

Sum of Squares

SSTR = 0.3116
SSTO = .6376

(a). Is this a fixed effect or random effect experiment? Explain why. Are the looms equal in the output? Use $\alpha = 0.05$.
(b). Estimate the variability between the looms.
(c). Estimate the experiment error variance.

6. Three different washing solutions are being compared to study their effectiveness in retarding bacterial growth in five-gallon milk containers. The analysis is done in a laboratory, and only three trials can be run on any day. Because days could represent a potential source of variability, the experimenter decides to use a randomized block design. Observations are taken for four days, and the data are shown here. Construct the ANOVA Table, and then analyze the data from this experiment (use a = 0.05) and draw conclusions.

	Days (B)				
Solution (A)	1	2	3	4	Row Mean
1	13	22	18	39	23
2	16	24	17	44	25.25
3	5	4	1	22	8
Column Mean	11.33	16.67	12	35	Grand Mean =18.75

7. Several Investigators describe an experiment to investigate the warping of copper plates. The two factors studied were the temperature (Factor A) and the copper content (Factor B) of the plates. The response variable was a measure of the amount of warping.

Temperature	Copper Content (%)			
	40	60	80	100
50	17, 20	16, 21	24, 22	28, 27
75	12, 9	18, 13	17, 12	27, 31
100	16, 12	18, 21	25, 23	30, 23
125	21, 17	23, 21	23, 22	29, 31

(a). Complete the ANOVA Table:

Source	S.S.	DF	MS	F
A	698			
B	156			
AB	114			
Error				
Total	1077			

(b). Is there any interaction between the factor?

(c). Is there any indication that either factor affects the amount of warping?

8. Consider a company that purchases its raw material from different suppliers. The company wishes to determine if the purity of the raw material is the same from each supplier. Three suppliers are chosen at random from a supplier list. There are 4 batches of raw material selected from each of 3 suppliers. Three measurements of purity are to be taken from each batch.

(a). Write down the correct effect model for the problem, and give the necessary assumptions.

(b). Specify the appropriate ANOVA table and how to test the hypothesis of the effects.

9. Four corn varieties were tested for their production in an experiment with four blocks. The following yield data were obtained: denote $\bar{y}_{i\cdot}$ and $\bar{y}_{\cdot j}$ the row mean w.r.t. corn types and column (block) mean of yield (y_{ij}) respectively, we have

$$\sum_{i=1}^{4} 4(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = 0.30, \sum_{j=1}^{4} 4(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2 = 0.70, \sum_{i=1}^{4}\sum_{j=1}^{4} 4(y_{ij} - \bar{y}_{\cdot\cdot})^2 = 1.11.$$

(a). Specify the design method and write down the appropriate effect model and restrictions to analyze the data.

(b). Test if the effect due to corn varieties is significant. [$F_{0.05}(3, 9) = 3.86$, $F_{0.05}(4, 16) = 3$]

(c). Following the model notations given in (a), find unbiased estimators for overall mean μ treatment effect α_i. Derive the distribution of the treatment effect based on the assumptions in (a).

10. Try to fit a simple linear regression model to the data on salt concentration (y) and the roadway area (x). Some calculation results:

$$\bar{x} = 0.824, \bar{y} = 17.1, \sum_{i=1}^{8}(x_i - \bar{x})^2 = 3.67, \sum_{j=1}^{8}(x_i - \bar{x})(y_i - \bar{y}) = 64.4, \sum_{j=1}^{8}(y_i - \bar{y})^2 = 1188.$$

(a). Write down the simple linear regression model and its necessary assumptions. Calculate the least squares estimates of the linear coefficients.
(b). Construct the ANOVA table for the regression model and estimate the error variance σ^2 . Find the coefficient of determination R^2 and explain it.

11. The Castle Bakery Company supplies wrapped Italian bread to a large number of supermarkets in a metropolitan area. An experimental study was made of the effects of heights of the shelf display (Factor A: bottom, middle, top) and the width of the shelf display (Factor B: regular, wide) on sales of this bakery's bread. Twelve supermarkets, similar in terms of sales volume and clientele, were utilized in the study. The six treatments were assigned at random to two stores.

Sales	Regular	Wide
Bottom	47, 43	46, 40
Middle	62, 68	67, 71
Top	41, 39	42, 46

(a). Specify the effect model and levels of the factors with necessary assumptions, and complete the ANOVA table given that SSA=1544, SSB=12, SSE=62, SSTO=1642.
(b). Is the interaction effect significant in the ANOVA analysis? Draw interaction plot and compare it with the F test outcome. [$F_{0.05}(1, 6) = 6.0$, $F_{0.05}(2, 6) = 5.14$, $F_{0.05}(3, 6) = 4.75$]
(c). Based on model assumptions, derive the expectations of Sum Square of Errors (SSE) and Sum Squares of Factor A (SSA), i.e. E(SSE) and E(SSA).

12. The effects of temperature (factor 1) and reaction time (factor 2) on percent yield of a certain chemical reaction (Response Y) are studied. There are two levels of temperature, 110°C and 130 °C, and two levels of reaction time, 50 minutes and 70 minutes. The 2^2 factorial experiment was replicated twice (n=2) and the order of the runs was randomized. The result is listed in the table:

	Design		Average
Run	x_1	x_2	Yield ($\bar{y}$)
1	–	–	55
2	+	–	60.6
3	–	+	64.2
4	+	+	68.2

(a). Estimate the main effects and the interaction effect.
(b). The pooled variance estimate is $s_p^2 = 0.375$. Find an approximate 95% confidence limits for main effects and interaction effect. What conclusions can you reach?

Brief Keys:

1. Observed χ_o^2 = 22.08, significant association.
2. (a). F = 0.86 , no significant difference between the variability.
   (b). pooled t-test t = -1.35, the flow rate does not affect the uniformity

3. (a) $\hat{\beta}_0 = -73.65, \hat{\beta}_1 = 0.516$ (b). Yes, it is a good linear fit with R^2 = 97%

4. (a) F=12.73, significant (b). 490 ± 174.5, so they are different.

5. (a) Random effect. F=4.78, the looms varies in the output.
   $(b).\hat{\sigma}_\tau^2 = 0.0123, (c).\hat{\sigma}^2 = 0.0163$

6.

Source	SS	DF	MS	F
A	703.5	2	351.75	40.7
B	1106.9	3	369	42.7
Error	51.8	6	8.64	
Total	1862.2	11		

7.

Source	SS	DF	MS	F
A	698	3	232.8	34.3
B	156	3	52	7.7
AB	114	9	12.6	1.86
Error	109	16	6.78	
Total	1077	31		

8.

(a) Effect model:

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{k(i,j)}, i = 1,...,a; j = 1,...,b; k = 1,...,n$$

Where factor A (supplier) is the main effect , and factor B (batch) are the nested factor within the supplier factor. a=3, b=4, n=3. Both factors are random. The assumption for the model is

$$\alpha_i \sim N(0, \sigma_\alpha^2), \beta_j \sim N(0, \sigma_\beta^2), \varepsilon_{k(i,j)} \sim N(0, \sigma^2),$$

and all effects are mutually independent.

(b).

Source	SS	DF	MS	F
A	SSA	2	MSA	F_A =MSA/MSB(A)
B(A)	SSB(A)	9	MSB(A)	$F_{B(A)}$=MAB(A)/MSE
Error	SSE	24	MSE	
Total	SSTO	35		

Use F_A to test main effect of factor A: $H_0 : \sigma_\alpha^2 = 0$

Use $F_{B(A)}$ to test the effect of nested factor B(A) : $H_0 : \sigma_\beta^2 = 0$

9. a). Randomized Complete Block Design

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{i,j}, i = 1,...,k;\ j = 1,...,b$$

b). F =8.18 > F(0.05, 3, 9)=3.86, reject $H_0 : \alpha_i = 0, i = 1,..., k$.

c). Overall mean $\hat{\mu} = \overline{Y}_{..}$ treatment effect $\hat{\alpha}_i = \overline{Y}_{i.} - \overline{Y}_{..}$ follows a normal distribution with mean 0 and variance $\left(\frac{k-1}{k} \cdot \frac{\sigma^2}{b}_{..} \right)$.

10. (a) $\hat{\beta}_0 = 2.64, \hat{\beta}_1 = 18.548$ (b). SSR=1130, MSE=9.67, R^2 = 96.8%

11. (a). 2-way factorial design with fixed factors

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, i = 1,..., a;\ j = 1,..., b; k = 1,..., n$$

where *i.i.d.* error $\varepsilon_{ijk} \sim N(0,\sigma^2)$ and a=3, b=2, n=2.

Source	SS	DF	MS	F
A	1544	2	772	75
B	12	1	12	1.17
AB	24	2	12	1.17
Error	62	6	10.3	
Total	1642	11		

(b). F_{AB} = 2.32 < F (0.05, 2, 6)

(c). $E(SSE) = ab(n-1)\sigma^2$, $E(SSA) = (a-1)\sigma^2$.

12. (a) Main effect: $(1) = 2.4, (2) = 4.1$, interaction effect: $(12) = -0.4$.

(b). Confidence interval for the effects: $2.4 \pm 0.43,\ 4.1 \pm 0.43,\ -0.4 \pm 0.43$.