Name: ____________________ Spring, 2017

Applied Statistics Comprehensive Examination

- Calculators are permitted on this examination.
- When you compute a confidence interval, always give an interpretation of the interval in the context of the problem.
- When you perform a hypothesis test, always write down the null and alternative hypotheses, and write the conclusion in the context of the problem.
- There are 200 points on this examination.
- You must give complete explanations to receive full credit.

1. (20 points) The following table gives the number of wins and losses for the Villanova Men's Basketball team over the four seasons from 2012-2016.

	2012-2013	2013-2014	2014-2015	2015-2016
# Wins	20	29	33	35
# Losses	14	5	3	5

Using level 0.05, test for a difference in the true yearly winning percentages.

2. (20 points) A simple linear regression was run on a set of data. You are given only the following information:

1. $\hat{y} = 11.5 - 1.5x$.
2. The t-test for $H_0 : \beta_1 = 0$ was not significant at $\alpha = .05$ level. A computed t of -4.087 was compared to $t_{(0.025,2)}$.
3. The estimate of σ^2 was 1.75.

(a) (15 points) Complete the analysis of variance table using the given results.

(b) (5 points) Compute and interpret the coefficient of determination R^2.

3. (20 points) A local archery club gives a free membership to anyone who can hit a certain target at least 30 out of 50 times at an open tryout.

(a) (7 points) If Archer A hits the target at a true rate of 55%, what is the approximate probability that Archer A wins a free membership?

(b) (7 points) If Archer B hits the target at a true rate of 70%, what is the approximate probability that Archer B does not win a free membership?

(c) (6 points) List any assumptions or conditions needed for the calculations in (a) and (b) and comment on whether they are reasonable.

4. (30 points) An experiment was conducted to compare the yields of five varieties of blight-resistant corn (A, B, C, D, E) using a balanced completely randomized design with 25 observations. The following summary information was obtained: Variety $MS = 598$, residual $SS = 3600$, and the variety means were 80, 85, 70, 82, and 93, respectively.

(a) (10 points) Complete the ANOVA table and test the equality of variety means using a 5% level of significance.

(b) (10 points) Compute a 95% confidence interval for $\tau_A - \tau_C$.

(c) (10 points) Perform Fisher's LSD procedure with $\alpha = 0.05$ and summarize the results using the underline method.

5. (40 points) A stock broker is choosing between two investment strategies, Strategy A and Strategy B. She runs an experiment in which 50 of the next 100 business days are randomly assigned to A, with the other 50 assigned to B. She obtains the following summary data for daily gains (losses if negative) in tens of thousands of dollars.

	$\overline{y}$	s	Minimum	1st Quartile	Median	3rd Quartile	Maximum
Strategy A	1.02	1.73	-1.60	-0.19	0.79	2.81	5.85
Strategy B	0.92	1.20	-0.80	0.17	0.75	2.34	4.08

(a) (10 points) Find and interpret a 95% confidence interval for the difference in mean daily gains for the two strategies.

(b) (5 points) Based on the confidence interval in (a), is there a difference between the mean daily gains for the two strategies? Write a one sentence explanation.

(c) (10 points) List any assumptions or conditions needed for the confidence interval in (a) and comment on whether they are reasonable.

(d) (10 points) The stock broker now decides that she prefers the strategy that has the lower variance in daily gains. Using level 0.05, test for evidence that the variance is lower for Strategy B than Strategy A.

(e) (5 points) List any assumptions or conditions needed for the hypothesis test in (d) and comment on whether they are reasonable.

6. (20 points) Consider an experiment with a two way treatment structure having three rows, two columns and four observations per treatment combination in a completely randomized design.

(a) (5 points) Specify the population marginal means (PPM) for all three rows assuming a cell means model.

(b) (5 points) Specify the population marginal means for both columns assuming an effects model with interaction (use α_i for row effects, β_j for column effects and γ_{ij} for interaction effects).

(c) (5 points) Specify two orthogonal contrasts for interaction effects assuming a cell means model.

(d) (5 points) Determine if $\alpha_1 - \alpha_2 + \frac{1}{2}(\gamma_{11} + \gamma_{12}) - \frac{1}{2}(\gamma_{21} + \gamma_{22})$ is estimable and justify your answer.

7. (20 points) In a statistics class, students were assigned the task of determining whether Villanova students own more than 15 pairs of shoes on average. One student in the class surveyed 100 students and found a sample mean of 16.5 pairs of shoes, with a sample standard deviation of 6.0 pairs.

(a) (10 points) Using this student's data, conduct the test of interest at the 0.10 level.

(b) (5 points) List any assumptions or conditions that are needed for the test in (a) and comment on whether they are reasonable.

(c) (5 points) Suppose that the true average number of pairs of shoes is 18. Estimate the sample size needed for the test to have a power of 90%.

8. (30 points) A study examined survey data from a number of hospitals to determine factors that are associated with the probability of acquiring an infection. Six independent variables were considered, descriptions of which are given in the table.

Variable Name	Description
Age	Average age of patients (in years)
Length	Average length of stay of all patients in hospital (in days)
Region	Geographic region (NE, NC, S, or W)
Nurses	Number of full-time-equivalent registered and licensed nurses
DailyPatient	Average number of patients in hospital per day during study period
MedSchool	Is the hospital affiliated with a medical school?(1=Yes; 0=No)

Assuming that all of the usual model assumptions are satisfied, refer to the model output on the next page to answer the following questions.

(a) (5 points) Find the number of surveyed hospitals, n.

(b) Refer to the output for Model 1.
(i) (5 points) Interpret the estimated coefficient $\hat{\beta}_{\text{RegionW}}$.
(ii) (10 points) Find a 90% confidence interval for β_{Length}.

(c) (10 points) Test whether the coefficients for being affiliated to a medical school, for the average number of patients, and for their interaction are all simultaneously zero, at $\alpha = 0.05$, while keeping the other predictor variables in the model.

See next page for output related to Problem 8.

Model 1: Coefficient summary for the full model with all the independent variables

	Coefficient	Standard Error	t Stat	P-value
Intercept	1.093	1.387	0.788	0.4325
Age	-0.025	0.024	-1.016	0.3120
Length	0.414	0.074	5.604	1.76e-07
RegionNC	0.124	0.299	0.415	0.6789
RegionS	-0.121	0.302	-0.402	0.6888
RegionW	0.823	0.389	2.117	0.0367
Nurses	0.004	0.002	1.983	0.0500
MedSchool	0.952	0.775	1.229	0.2217
DailyPatient	-0.001	0.002	-0.266	0.7906
MedSchool*DailyPatient	-0.004	0.002	-1.839	0.0688

Analysis of Variance

Model 1: Risk $\sim$ Age + Length + Region + Nurses + MedSchool + DailyPatient + MedSchool*DailyPatient

Model 2: Risk $\sim$ Age + Length + Region + Nurses

Model	Residual sum of squares	Residual degrees of freedom
Model 1	117.9	103
Model 2	124.8	106