IE 330 Seat # ________ Name ___ < KEY > ___

Open book and notes. 120 minutes.

Covers Chapters 8 through 14 of Montgomery and Runger (fourth edition).

Cover page and eight pages of exam.
No calculator.

(2 points) I have, or will, complete a course evaluation.

____
...sign here...

NEITK: "not enough information to know"

iid: "independent and identically distributed"

An element of an experiment is "random" if its value can change with each replication of the experiment.

A bar over notation, such as $\bar{X}$, denotes the sample mean.

There is no need to simplify calculations.

Score ________________________

Open book and notes. No calculator. 120 minutes.

1. True or false. (2 points each)

(a) T F ← In hypothesis testing, failing to reject H_0 leads to the strong claim that H_0 is true.

(b) T F ← In regression analysis, consider a fitted model that passes through all data points. Because the fit cannot be better, using the fitted model for interpolation also has no error.

(c) T ← F Confidence-interval length goes to zero as sample size goes to infinity.

(d) T F ← Student's T distribution is used to compare variances.

(e) T ← F ← If two factors have an interaction effect, the effects are additive.

(f) T ← F In hypothesis testing, with everything else being unchanged, more power is good.

(g) T F ← If only two factors are confounded, doubling the number of observations at each design point will "unconfound" them.

(h) T ← F In hypothesis testing, the greater the distance between the values in H_0 and H_1, the greater the power of the test.

(i) T F ← For all normal distributions, a bit more than 75% of the probability is within one standard deviation of the mean.

(j) T ← F If (a, b) is a 95% confidence interval for the variance σ^2, then $(\sqrt{a}, \sqrt{b})$ is a 95% confidence interval for the standard deviation σ.

(k) T ← F If a sample is iid, then $\sigma^2 = n\ \mathrm{V}(\bar{X})$, where σ denotes the standard deviation of the n observations X.

2. (Montgomery and Runger, Problem 11–1) An article in *Concrete Research* ("Near Surface Characteristics of Concrete: Intrinsic Permeability," 41, 1989), presented data on comprehensive strength x and intrinsic permeability y of various concrete mixes and cures. Summary quantities are $n = 14$, $\sum y_i = 572$, $\sum y_i^2 = 23{,}530$, $\sum x_i = 43$, $\sum x_i^2 = 157.42$ and $\sum x_i y_i = 1697.80$. Consider the regression model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$.

(a) (4 pts) Write the equation for the fitted model's slope, which is also the estimated value of dy / dx, the first derivative of the expected response with respect to the compressive strength.

The formula for $\hat{\beta}_1$. ←

(b) (3 pts) Suppose that $\hat{\beta}_0$ and $\hat{\beta}_1$ have been calculated. Suppose that the observed value of permeability at $x = 3.9$ is $y = 50.1$. State the corresponding residual.

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \varepsilon_i$$

implies

$$\varepsilon_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

implies

$$\varepsilon_i = 50.1 - (\hat{\beta}_0 + 3.9\hat{\beta}_1) \leftarrow$$

---

3. A manufacturer of semiconductor devices takes a random sample of size n of chips and tests them, classifying each chip as defective or nondefective. Let $X_i = 0$ if the chip is nondefective and $X_i = 1$ if defective. The sample fraction defective is $\hat{p} = (X_1 + X_2 + \cdots + X_n) / n$. Let p denote the fraction of all chips that are defective.

(a) (3 pts) We take a sample from a "population". What is the population in this example?

The population is all chips made by this manufacturer. ←

(b) (3 pts) What does the Central Limit Theorem say about the distribution of $\hat{p}$?

In the limit as n goes to infinity, $\hat{p}$ is normally distributed.

(c) (3 pts) State the expected value of $\hat{p}$.

---

p ←

(d) (3 pts) State the standard error of $\hat{p}$.

$[p\,(1 - p) / n]^{1/2}$ ←

---

4. (Montgomery and Runger, fourth edition, Problem 8–6) Following are two confidence intervals of the mean μ of the cycles-to-failure of an automobile door-latch mechanism. Both intervals are calculated from the same sample data.

$31{,}249 \le \mu \le 32{,}157$ ← and $31{,}105 \le \mu \le 32{,}301$

(a) (3 pts) What is the value of the sample mean of cycles-to-failure?

For any α value, the sample mean is the center of interval. Therefore,

$$\frac{31{,}249 + 32{,}157}{2} = \frac{31{,}105 + 32{,}301}{2} = 31{,}703 \leftarrow$$

(b) (2 pts) The confidence level for one of these confidence intervals is 95%; the other is 99%. Circle the 95% confidence interval. → Circle the left interval. ←

(c) (3 pts) Explain why your choice in Part (b) is correct.

The 95% interval covers less often, so it is shorter. ←

(d) (3 pts) In the expression

$$P(31{,}249 \le \mu \le 32{,}157) = 1 - \alpha,$$

which quantities are random?

After the observations are obtained, nothing is random. ←

or

The constants "31,249" and "32,157" are observations of random variables. ←

---

5. Consider Example 14–2 on pages 556–557 of Montgomery and Runger (fourth edition). Surface roughness is studied as a function of feed rate, depth of cut, and tool angle. Data are given in Table 14–10; an ANOVA is given in Table 14–11.

(a) (3 pts) For the feed rate, $F = 18.69$ and $P = 0.003$. If the F value had been greater, then p would have been

(choose one) (i) smaller ← (ii) greater (iii) unchanged (iv) NEITK

(b) (3 pts) Again consider feed rate. The value $P = 0.003$ is small, in the sense that most analysts would say that the null hypothesis H_0 is rejected in favor of the alternative hypotheses H_1. State H_0 and H_1.

H_0: feed rate does not affect mean roughness.

H_1: feed rate does affect mean roughness.

(c) (3 pts) The two feed rates are 20 and 30 inches per minute; feed rate is found to be significant. Suppose that instead the experiment had been run with feed rates 20 and 25 inches per minute. Would the feed-rate P value likely become larger ←, smaller, or remain the same? (circle one)

(d) (3 pts) There are sixteen observations and fifteen degrees of freedom. The "lost" degree of freedom is because μ, the first term in the linear model, is estimated from the data. But μ is not in the ANOVA table. In words, state the meaning of μ.

μ is the grand mean, over all 16 design points. ←

(e) (3 pts) For this and all other ANOVA tables, what is the smallest possible F value?

---

zero ←

---

6. Consider a two-factor fixed-effects factorial model: $Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$ for $i = 1, 2, ..., a$, $j = 1, 2, ..., b$, and $k = 1, 2, ..., n$. Call the two factors A and B.

(a) (2 pts) How many observations are in this experiment? ___ < abn > ___

(b) (3 pts) The observations should be "run" in a random order. Why?

To reduce the any effects due to time order. ←

(c) (2 pts) The total sum of squares is

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{abn}.$$

Carefully place parentheses to make the formula unambiguous.

$$SS_T = \left[\sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (y_{ijk}^2) \right] - \left[\frac{y_{...}^2}{abn} \right].$$

(d) (3 pts) The mean square due to error is $MS_E = SS_E / [ab(n-1)]$. It appears in the denominator of the F statistics. What does MS_E estimate?

The variance of ε_{ijk},
which is assumed to be equal for all ijk. ←

7. Consider the chi-squared statistic $(n-1)\, S^2 / \sigma^2$.

(a) (2 pts) How many populations are represented in the definition of the chi-squared statistic?

---

1 ←

(b) (3 pts) Which part, or parts, of $(n-1)\, S^2 / \sigma^2$ are random?

S^2 ←

(c) (4 pts) Consider a random sample from a population with variance σ^2. For a sample of size $n = 3$, use Table IV to sketch the cumulative distribution function. Label and scale both axes.

Sketch a horizontal and a vertical axis.
Label the horizontal axis with any dummy variable, such as x.
Label the vertical axis with $F(x)$.
Scale the vertical axis with zero and one.
$n = 3$ corresponds to $v = 2$ degrees of freedom in Table 4.
Plot some points for various values of α
Sketch a smooth curve through the points.
Scale the horizontal axis with at least two numbers.

(d) (3 pts) This statistic is useful for

(i) confidence intervals   (ii) hypothesis testing   (iii) both ←

8. Montgomery and Runger (Problem 10–31) The manager of a fleet of automobiles is comparing two brands of radial tires. She assigns one tire of each brand at random to the two rear wheels of eight cars and runs the cars until the tires wear out. The data (in kilometers) follow.

Car	1	2	3	4	5	6	7	8
Brand 1	36,925	45,300	36,240	32,100	37,210	48,360	38,200	33,500
Brand 2	34,318	42,280	35,500	31,950	38,015	47,800	37,810	33,215

Generically, let X_{i1}, X_{i2} denote the results from the i th car.

(a) (3 pts) Estimate the mean difference between brands, $\mu_1 - \mu_2$.

$\bar{X}_1 - \bar{X}_2$ or $\bar{D}$, where $D_i = X_{i1} - X_{i2}$.

(b) (3 pts) Explain how we know that the manager plans a Paired-T test (rather than a regular T test).

She assigned one tire of each brand to each car. ←

(c) (2 pts) T F ← The Paired-T test assumes that X_{i1} and X_{i2} are independent.

(d) (3 pts) The manager wants to buy Brand 1 tires. Therefore, she plans to hide the results if Brand 2 wins. She wants to report that she did the experiment carefully, with a probability of type I error of $\alpha = 0.05$. Given her motive, should she perform a one-sided, or two-sided, test?

one sided ←