ENGI 4421 Final Examination

2019 August

1. A quality control process includes the selection of a random sample of five items from a batch of 100. The items in the sample are tested to destruction. [10]

   Find the number of distinct ways the random sample could be selected.
   Show your working.

---

2. Measurements of diameter and length are taken for a sample of 300 cables. The results are categorized in the contingency table below. [12]

			Length		
	Observed	***short***	***medium***	***long***	**Total**
Diameter	***narrow***	11	10	9	**30**
	medium	20	46	54	**120**
	wide	9	30	51	**90**
	very wide	10	14	36	**60**
	Total	**50**	**100**	**150**	**300**

   Use a chi-square test to determine whether the diameter and length are independent.

---

3. The efficiency x (in appropriate units) of 20 randomly selected motors performing certain tasks is measured. A new controller system is installed on those 20 motors and the efficiency y is measured again. An agent claims that the new controller system increases the efficiency by at least 2.0. It is known that both populations are normally distributed.

(a) Which of the two sample t-tests (paired or unpaired) should be conducted? State the reason for your selection. [3]

(b) Conduct the appropriate hypothesis test, at a level of significance of 1%, using whichever one of these two sets of Minitab® output is appropriate: [6]

Paired T-Test and CI: x, y

Estimation for Paired Difference

Mean	StDev	SE Mean	99% Upper Bound for µ_difference
-3.261	2.055	0.460	-2.094

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
x	20	76.50	5.92	1.32
y	20	79.76	6.49	1.45

Test

Null hypothesis H_0: µ_difference = -2

Alternative hypothesis H_1: µ_difference < -2

T-Value	P-Value
-2.74	0.006

Two-Sample T-Test and CI

Method

μ_1: mean of x

μ_2: mean of y

Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Difference	99% Upper Bound for Difference
-3.26	1.51

Test

Null hypothesis H_0: $\mu_1 - \mu_2 = -2$

Alternative hypothesis H_1: $\mu_1 - \mu_2 < -2$

T-Value	DF	P-Value
-0.64	37	0.262

(c) Does the evidence support the agent's claim? [2]

4. It is known that each new cladding panel manufactured by an experimental new process has a failure probability of 0.05, independently of all other panels. A random sample of six panels is taken. Let the random quantity X represent the number of defective panels in the random sample.

(a) Is the probability distribution of X exactly binomial, approximately binomial or not binomial at all? [2]

(b) Find $\mathrm{E}[X]$ and $\mathrm{V}[X]$. [2]

(c) Find, correct to three significant figures, the probability that there are no defective panels in the random sample. [3]

(d) Now suppose that random samples of size 6 are drawn from each of 100 batches. Let $\overline{X}$ be the mean value of X across these 100 samples. Find $\mathrm{P}\left[\overline{X} > 0.4\right]$. You do **not** have to use linear interpolation. [5]

---

5. The mass M of a shipping box of bolts is known to be a random quantity that follows a Normal distribution. It is believed that the population mean mass is 100 kg and the strength of that belief is represented by the standard deviation $\sigma_o = 5\,\mathrm{kg}$. A random sample of 51 such boxes has a mean mass of 94.8 kg with a sample standard deviation of 21.2 kg.

(a) Construct a classical 95% confidence interval estimate for the true mean mass μ. [5]

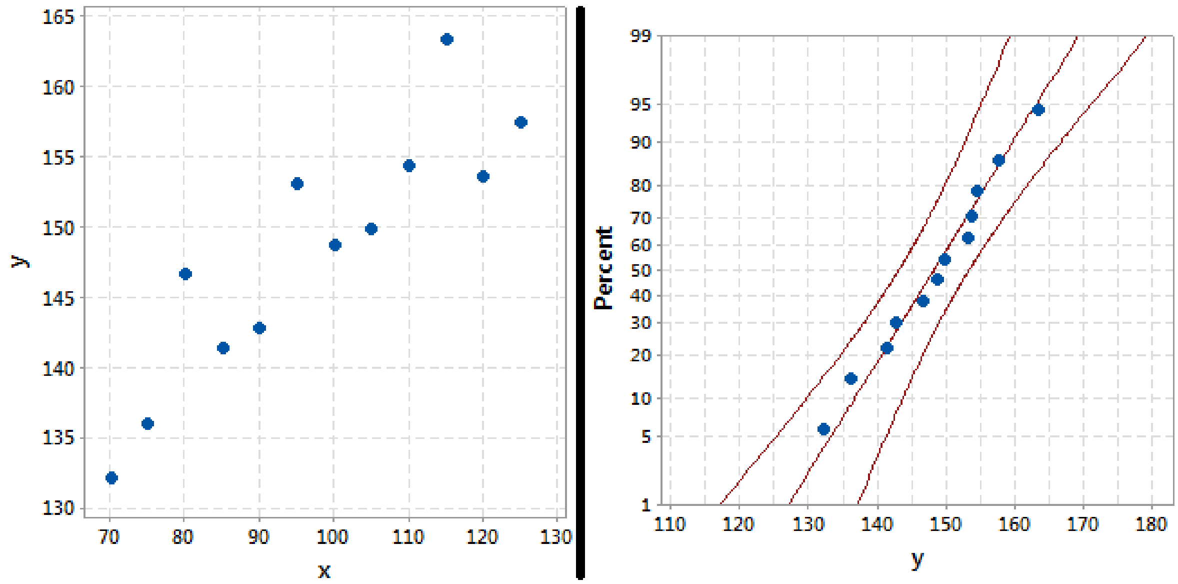
(b) Construct a Bayesian 95% confidence interval estimate for the true mean mass μ. [5]

(c) Is there sufficient evidence to conclude that the true mean mass μ is not 100 kg? [2]

(d) Provide a *brief* reason for the different widths of your two confidence intervals. [2]

☞

6. For 12 observations of temperature (x), the corrosion rate (y) of a metal beam is measured. A scatterplot and a normal probability plot of the observations are presented here:



(a) State ***two*** reasons why it is appropriate to use the simple linear regression model. [2]

(b) Given the summary statistics [4]

$$n = 12 \qquad \sum x = 1170.00 \qquad \sum y = 1778.80$$
$$\sum x^2 = 117\,650.00 \qquad \sum xy = 175\,022.00 \qquad \sum y^2 = 264\,572.15$$
$$n\,S_{xx} = 42\,900.00 \qquad n\,S_{xy} = 19\,068.00 \qquad n\,S_{yy} = 10\,736.36$$

find the equation of the regression line $y = \hat{\beta}_0 + \hat{\beta}_1 x$.

(c) Construct the ANOVA table from these data. [6]

(d) To the nearest one per cent, what proportion of the variation in Y is explained by the linear regression? [2]

(e) Find the 95% prediction interval for a future observation of Y at $x = 97.5$. [5]

7. A manager knows that 75% of all crates of components are "good" (that is, contain no defective components), while the other 25% of crates are "bad" (that is, contain at least one defective component).

A test is available, costing \$5 per crate, to determine whether a crate contains any defective components. A "bad" crate is absolutely certain to fail the test. However the test is not perfect. A "good" crate has a probability of 0.2 to fail the test.

Accepting a good crate earns \$30.
Accepting a bad crate causes a loss of \$60.
Rejecting a crate causes a loss of \$10 (regardless of whether it is good or bad).

(a) Show that the probability that a crate is good, given that it failed the test, is $\frac{3}{8}$. [5]

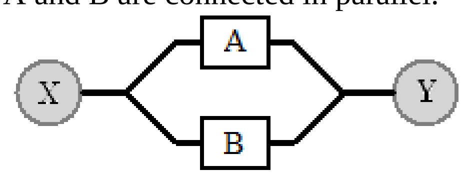
(b) Find the odds corresponding to the probability $p = \frac{3}{8}$. [2]

(c) Draw a decision tree to illustrate this situation and determine the optimum strategy that maximizes the expected gain per crate for the contractor and find the exact value of that maximum expected gain. [15]

---

8. *BONUS QUESTION* [+5]

Two pumping stations A and B are connected in parallel.



Let A = the event that station A is working
and B = the event that station B is working
It is known that $\mathrm{P}[A \mid B] = 0.90$, $\mathrm{P}\left[A \mid \tilde{B}\right] = 0.75$ and $\mathrm{P}[B \mid A] = 0.50$

Find $\mathrm{P}[A \cup B]$ (the probability that fluid flows from X to Y).

☜ Back to the index of questions On to the solutions ☞