Stevens Institute of Technology

SYS-601 Practice Exam #2 Solutions

2.1 Café Quality Control

To improve the quality of coffee products, Café Java has implemented a data collection and analysis program to measure the temperature of all coffee products. The attached file `cafe_temps.csv` includes measurements for a random sample of 10 cups of coffee.

(a) Which probability distribution models the mean of a small number of samples?

(A) χ^2 (B) F (C) X t (D) z

Using this distribution, compute a 99% confidence interval for mean coffee temperature.

$$\bar{x} \pm t_{0.005} \frac{s}{\sqrt{N}} = 172.96 \pm 3.25 \frac{4.09}{\sqrt{10}} = [\quad 168.8 \quad , \quad 177.2 \quad]$$

Use a hypothesis test to check whether the data support Café Java's goal to achieve a mean coffee temperature of precisely 175°F. Report the test statistic and p-value.

$$t_{N-1} = \frac{\bar{x} - 175}{s/\sqrt{N}} = \frac{172.96 - 175}{4.09/\sqrt{10}} = -1.58 \qquad p = 2 \cdot (1 - F_{\text{tdist}}(t_{N-1})) = 0.149$$

(b) Based on the above distribution, how many cups of coffee per million opportunities are expected to be *below* a minimum temperature of 165°F and worthy of a refund?

$$t_{N-1} = \frac{165 - \bar{x}}{s/\sqrt{N}} = \frac{175 - 172.96}{4.09/\sqrt{10}} = -6.15$$

$$N = 10^6 \cdot F_{\text{tdist}}(t_{N-1}, N-1) = 83.9 \text{ per million}$$

(c) Which probability distribution models the variance of a small number of samples assuming the population is normally distributed?

(A) X χ^2 (B) F (C) t (D) z

Use a hypothesis test to check whether the data support Café Java's goal to achieve a standard deviation in coffee temperature below 3.0°F. Report the test statistic and p-value.

$$\chi^2_{N-1} = \frac{(N-1) \cdot s^2}{\sigma^2} = \frac{9 \cdot 4.09^2}{3.0^2} = 16.7 \qquad p = 1 - F_{\text{chi2}}(\chi^2_{N-1}) = 0.053$$

2.2 Café Service Times

For a term project to model operations of Café Java, you observe $N = 50$ orders and measure how long it takes to prepare and complete the order in minutes (`cafe_times.csv`).

(a) What is $\bar{x}$, the average service time (minutes per order) observed in this sample?

$$\bar{x} = 0.52$$

(b) Recognizing that the sample mean $\bar{X}$ is a random variable based on the particular samples observed, what is the closest approximate distribution of $\bar{X}$?

(A) $\bar{X} \sim \text{binomial}(n = a_1, p = a_2)$

(B) $\bar{X} \sim \text{exponential}(\lambda = a_1)$

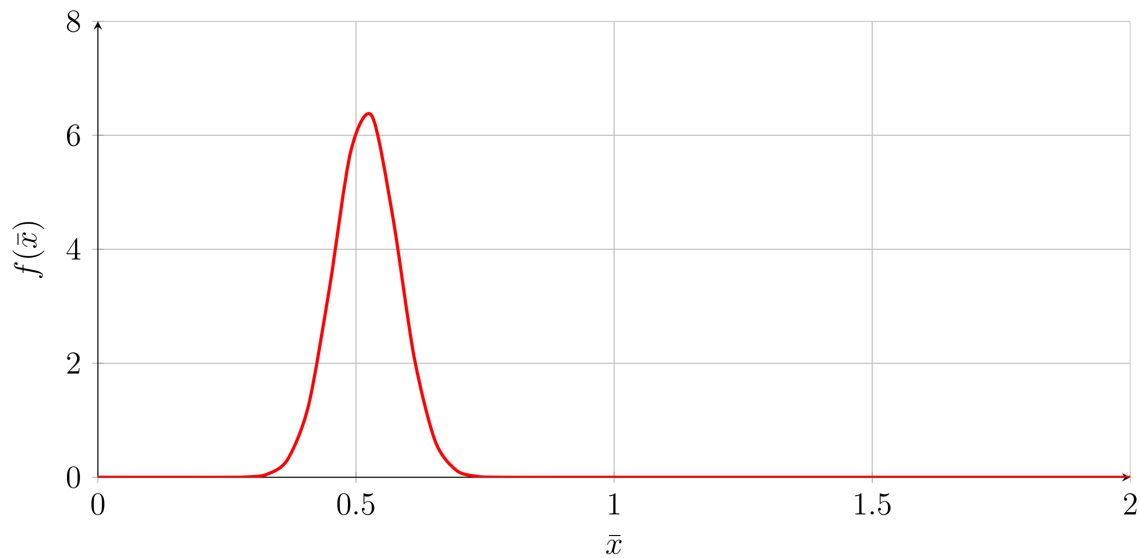
(C) X $\bar{X} \sim \text{normal}(\mu = a_1, \sigma = a_2)$

(D) $\bar{X} \sim \text{poisson}(\lambda = a_1)$

(c) What are the distribution parameter value(s) for the selected case above?

$$a_1 = \bar{x} = 0.52 \qquad a_2 = \frac{s}{\sqrt{N}} = \frac{0.44}{\sqrt{50}} = 0.062 \text{ (if required)}$$

(d) Create a PDF plot of the distribution selected above $f(\bar{x})$ for values $0 \leq \bar{x} \leq 2$.



(e) What is a 95% confidence interval for the population mean of $\bar{X}$?

$$\bar{X} \in \bar{x} \pm z_{0.025}\frac{s}{\sqrt{N}} = 0.52 \pm 1.96\frac{0.44}{\sqrt{50}} = [\quad 0.40 \quad , \quad 0.64 \quad]$$

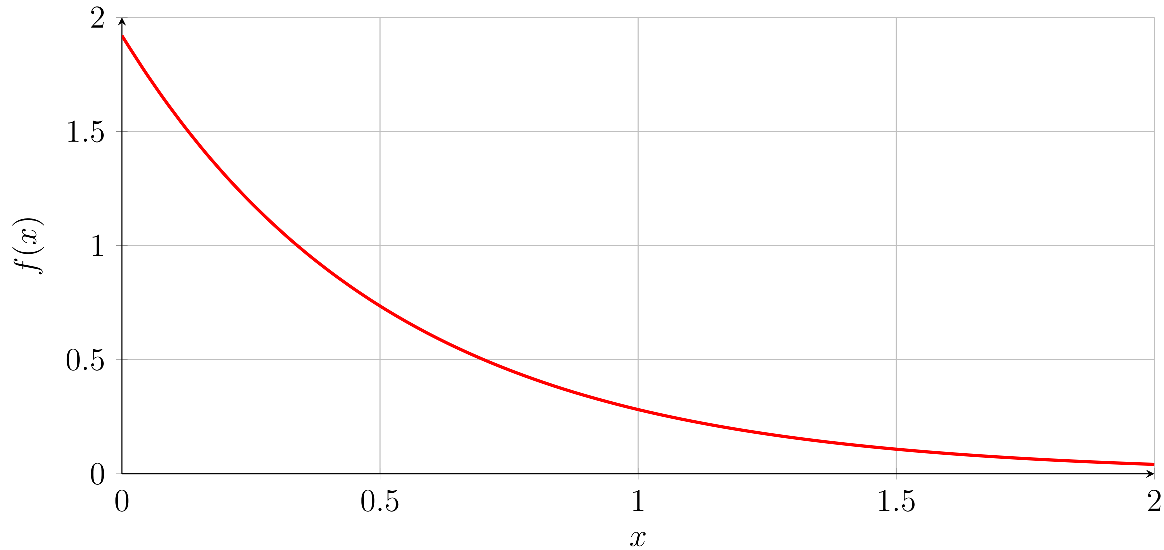
2.3 Café Service Modeling

Continuing with the dataset from the previous problem, you want to model the service times as a exponential distribution, i.e. $X \sim \text{exponential}(\lambda)$.

(a) Estimate the service rate parameter λ in units of **orders per minute**.

$$\lambda = \frac{1}{\bar{x}} = \frac{1}{0.52} = 1.92$$

(b) Create a PDF plot of the distribution selected above $f(x)$ for values $0 \leq x \leq 2$.



(c) Count the number of observations for each bin below.

x Bin:	$[0, 0.25)$	$[0.25, 0.5)$	$[0.5, 1.0)$	$[1.0, 2.0)$
Observed:	14	16	13	7

(d) Compute the expected number of observations in $N = 50$ samples assuming the converted probability mass values provided below.

x Bin:	$[0, 0.25)$	$[0.25, 0.5)$	$[0.5, 1.0)$	$[1.0, 2.0)$
$p(x) = \int f(x)dx$:	0.381	0.236	0.236	0.125
Expected:	$50 \cdot 0.381 = 19.05$	$50 \cdot 0.236 = 11.8$	11.8	6.25

(e) Perform a statistical test with a null hypothesis that the service times follow an exponential distribution and report the resulting test statistic and p-value.

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i} = 1.34 + 1.49 + 0.12 + 0.09 = 3.05 \qquad p = 1 - F_{\text{chi2}}(\chi^2_{4-1-1}) = 0.22$$

2.4 Café Employee Analysis

To study employee efficiency at Café Java, you want to perform an initial analysis to compare relative service rates for five servers: Dawn, Peter, David, Anthony, and Patrick. You know that espresso orders usually take longer to prepare than drip coffee orders. Understanding that ANOVA requires samples from approximately normal populations (rather than the exponential distributions hypothesized in the previous problem), you collect data on the number of orders completed in a five minute period (`cafe_employees.csv`).

(a) Calculate the following values for a two-way ANOVA:

$$SSC = 75.68 \qquad MSC = 18.92$$
$$SSI = 56.48 \qquad MSI = 14.12$$
$$SSE = 274.4 \qquad MSE = 6.86$$

(b) Perform a statistical test with a null hypothesis that all five employees have identical service rates and report the test statistic and p-value.

$$F = \frac{MSC}{MSE} = \frac{18.92}{6.86} = 2.76 \qquad p = 1 - F_{\text{fdist}}(F_{4,40}) = 0.04$$

Interpret this result in a format suitable for someone unfamiliar with statistics.

Results indicate at least one employee has a different service rate than the others. Statistical significance indicates there is only a 4% chance this statement is incorrect.

(c) Perform a statistical test with a null hypothesis that there is no interaction between employee and coffee type and report the test statistic and p-value.

$$F = \frac{MSI}{MSE} = \frac{14.12}{6.86} = 2.06 \qquad p = 1 - F_{\text{fdist}}(F_{4,40}) = 0.10$$

Interpret this result in a format suitable for someone unfamiliar with statistics.

Results are suggestive but not statistically significant that there is a non-zero interaction between employees and order on service rate. Statistical significance indicates there is a 10% chance this statement is incorrect.

2.5 Braking Model

The objective of this problem is to quantitatively model the stopping distance of a car as a function of its speed. The attached file `cars.csv` contains 50 experimental samples with the following fields:

- `speed` – initial speed when obstacle enters view (miles per hour)
- `dist` – stopping distance (feet)

(a) Develop a simple linear regression model. Report the following:

(i) Regression equation: $\hat{y} = \beta_0 + \beta_1 x$

(ii) Coefficient values and statistical significance:

Coefficient	Value	Statistically significant?
β_0	-17.58	X Yes No
β_1	3.93	X Yes No

(iii) Coefficient of determination: $r^2 = 0.651$

(b) Perform the following residual analysis for the `speed` factor:

(i) Are residuals homoscedastic? Yes X No

(ii) Are residuals independent? X Yes No

(c) An alternative regression model applies a natural log (ln) transform to both `speed` and `dist` to create a power law relationship. Report the following:

(i) Regression equation: $\ln \hat{y} = \beta_0 + \beta_1 \ln x \implies y = e^{\beta_0 + \beta_1 \ln x}$

(ii) Coefficient values and statistical significance:

Coefficient	Value	Statistically significant?
β_0	-0.73	X Yes No
β_1	1.60	X Yes No

(iii) Coefficient of determination: $r^2 = 0.73$

(d) Perform the following residual analysis for the transformed `speed` factor:

(i) Are residuals homoscedastic? X Yes No

(ii) Are residuals independent? X Yes No

(e) What is the predicted stopping distance for a car traveling 15 mph?

$$y = e^{-0.73+1.60 \cdot 15} = 37.0$$

(f) What is the predicted stopping distance for a car traveling 100 mph?

$$y = e^{-0.73+1.60 \cdot 100} = 772.5$$