Statistical Theory II

Practical Exam

5^{th} of May, 2026 | 9:00 – 12:00 | Auditorium 302

Instructions.

- Solve all problems in a single R script. Save the file as `FirstnameLastname.R` (e.g. `JohnSmith.R`).
- Begin your script with a comment containing your name and student ID: `# JohnSmith Student ID: ...`
- Where an answer requires interpretation or explanation, write it as a comment inside the script (use #).
- You may use `?` or `help()` for R documentation. **Internet, phones, and AI tools are strictly prohibited.**
- One A4 double-sided sheet of handwritten notes is permitted.
- Submit the final `.R` file via Moodle. If Moodle fails, e-mail it to paulius.kazlauskas@evaf.vu.lt.

Total: 10 points. Duration: 3 hours.

Problem 1. Bookstore Daily Orders *[2 points]*

A small online bookstore receives orders each day. The number of orders per day follows a Poisson(μ) distribution. Before looking at any data, the owner believes that μ has mean 9 and standard deviation 3.

Over the past 7 days, the daily order counts were:

$$8, \quad 12, \quad 7, \quad 11, \quad 9, \quad 13, \quad 10$$

(1) **[0.4 pts]** Determine the Gamma(r, v) prior distribution that matches the owner's belief. State r and v.

(2) **[0.4 pts]** Find the posterior distribution of μ. State r_1 and v_1.

(3) **[0.4 pts]** Compute the posterior mean, variance, and standard deviation. Verify numerically that the posterior mean equals the weighted average of the prior mean and the data mean.

(4) **[0.4 pts]** Compute the 95% credible interval for μ using the exact Gamma quantile function.

(5) **[0.4 pts]** Find the probability that the true daily order rate exceeds 11.

Problem 2. Mobile App Conversion Rate *[2 points]*

A startup launching a freemium mobile app wants to estimate π, the proportion of trial users who upgrade to the paid version. Based on industry benchmarks, the team considers three plausible values, $\pi \in \{0.3, 0.5, 0.7\}$, with the following prior beliefs:

π	Prior $P(\pi)$
0.3	0.2
0.5	0.5
0.7	0.3

Stage 1. A first batch of 10 trial users is observed; 3 of them upgrade.

(1) [**0.4 pts**] Compute the likelihood of the Stage 1 data for each value of π using `dbinom`. Then compute the posterior distribution of π given the Stage 1 data.

Stage 2. A second, independent batch of 10 trial users is observed; 4 of them upgrade.

(2) [**0.6 pts**] Update the posterior again, using your Stage 1 posterior as the new prior. Compute the new likelihood and the Stage 2 posterior.

(3) [**0.5 pts**] Repeat the analysis using the *single-step* approach: starting from the original prior, treat all 20 trial users as one observation (7 upgrades out of 20) and compute the posterior in a single update.

(4) [**0.5 pts**] Compare the posterior distributions you obtained in parts (2) and (3). Verify numerically that they are equal, and explain in one comment why this must be the case.

Problem 3. Beta Distribution Properties *[1.5 points]*

Let $Y \sim \text{Beta}(\alpha, \beta)$ with $\alpha = 8$ and $\beta = 4$.

(1) [**0.3 pts**] Compute $E(Y)$, $\text{Var}(Y)$, and $\text{SD}(Y)$ using the moment formulas of the Beta distribution.

(2) [**0.4 pts**] Compute $P(Y > 0.7)$ using the exact Beta cumulative distribution function.

(3) [**0.4 pts**] Compute the 5th and 95th percentiles of Y.

(4) [**0.4 pts**] Approximate $Y \approx N(E(Y), \text{Var}(Y))$ and recompute $P(Y > 0.7)$ using the normal approximation. Report the absolute difference between the exact and approximated values.

Problem 4. Defect Rate Estimation *[2 points]*

A factory inspector wants to estimate π, the proportion of items in a production batch that are defective. Based on records from previous batches, she believes that π has mean 0.2 and standard deviation 0.1.

A random sample of 100 items is inspected, and 18 of them are found to be defective.

(1) [**0.4 pts**] Determine the Beta(a, b) prior distribution that matches the inspector's belief. State a and b.

(2) [**0.4 pts**] Find the posterior distribution of π. State a_1 and b_1.

(3) [**0.4 pts**] Compute the posterior mean, variance, and standard deviation. Verify numerically that the posterior mean equals the weighted average of the prior mean and the data mean.

(4) [**0.4 pts**] Compute the 95% credible interval for π using the exact Beta quantile function.

(5) [**0.4 pts**] Find the probability that the true defect rate exceeds 0.25.

Problem 5. Used Car Prices *[2.5 points]*

The file `used_cars.xlsx`, available alongside this exam on Moodle, contains data on 150 used cars sold by a small dealership. The dataset has three columns:

- `price` — sale price in euros
- `mileage` — total kilometres driven
- `age` — age of the car in years

Loading the data. Save `used_cars.xlsx` in your working folder. To load it into R, use:

```
install.packages("readxl")
library(readxl)
cars <- read_excel("used_cars.xlsx")
```

(1) [**0.5 pts**] Load the dataset and briefly describe how `price` relates to `mileage`. You may use any approach you find appropriate — for example, summary statistics, a scatter plot, or the correlation between the variables. Write your description as a comment.

(2) [**0.7 pts**] Fit a simple linear regression of `price` on `mileage`. Report the estimated intercept and slope. Interpret the slope as a comment: what does the model predict happens to a car's price when its mileage increases by 1,000 km?

(3) [**0.8 pts**] Fit a second regression that adds `age` as a predictor (so the model now predicts `price` from both `mileage` and `age`). Compare the two models: does the adjusted R^2 improve? Does the coefficient on `mileage` change noticeably when `age` is added? In a comment, briefly explain what this tells you about the relationship between `mileage` and `age` in the data.

(4) [**0.5 pts**] Using your simple regression model from part (2), show how the fitted model compares to the observed data. The most natural way is a scatter plot of `price` against `mileage` with the fitted regression line overlaid (use `predict()` together with `lines()`, or `abline(model)`) — but other reasonable approaches comparing actual and fitted values are also acceptable.

End of exam. Total: 10 points.