Name	Matriculation number	TUM ID
Sama Elbaroudy	03768259	ge83muj
Sarah Berbuir	03767700	ge83cak
Mohamed Elbushnaq	03786474	go56cuh
Marina Aoki	03773384	ge94naz

Exercise 1

Informative Censoring:

- **Clinical Trials**: If participants drop out because they are experiencing severe adverse effects caused by the studied new treatment, this could be informative censoring. Their survival time might be shorter due to the treatment, providing valuable information about the treatment's efficacy and safety.
- **Epidemiological Studies**: In a study examining the survival rates of individuals with a specific infectious disease in a particular region, if some patients relocate due to access to better healthcare facilities or higher socioeconomic status, their censoring could be informative.

Non-Informative Censoring:

- **Geographical Relocation**: If some participants move out of the study area for reasons unrelated to their health status (e.g., job relocation, family reasons), their censoring would likely be non-informative.
- **Loss to Follow-up**: Participants might drop out due to reasons such as loss of contact, or unwillingness to continue participation. Their censoring is typically considered non-informative.

Exercise 2

1. $f(t) = -\frac{d}{dt} S(t)$

$\rightarrow$ $S(t) = P(T > t)$ definition of survival fn

$\frac{dS(t)}{dt} = \frac{d}{dt} P(T>t)$ differentiate both sides

$= -\frac{d}{dt} P(T \leq t)$ $\Rightarrow$ $P(T>t) = 1 - P(T \leq t)$

$\therefore -\frac{dS(t)}{dt} = f(t)$

2. $H(t) = -\log S(t)$

$h(t) = \frac{f(t)}{S(t)}$ $\rightarrow$ hazard fn

$= \frac{-\frac{dS(t)}{dt}}{S(t)}$ from 1.

$h(t) = -\frac{d}{dt}(\log S(t))$

$\int h(t)\,dt = \int -\frac{d}{dt}(\log S(t))\,dt$

$H(t) = -\log S(t) + C$

@ $t=0$ $H(t)=0$ & $S(t) = 1$ & $\log 1 = 0$

$\therefore C = 0$ $H(t) = -\log S(t)$

Exercise 3

$$S_1(t) = [S_0(t)]^R$$

$$\log S_1(t) = R \log S_0(t) \qquad \text{(log both sides)}$$

$$\frac{d}{dt} \log S_1(t) = R \frac{d}{dt} \log S_0(t) \qquad \text{(differentiate both sides)}$$

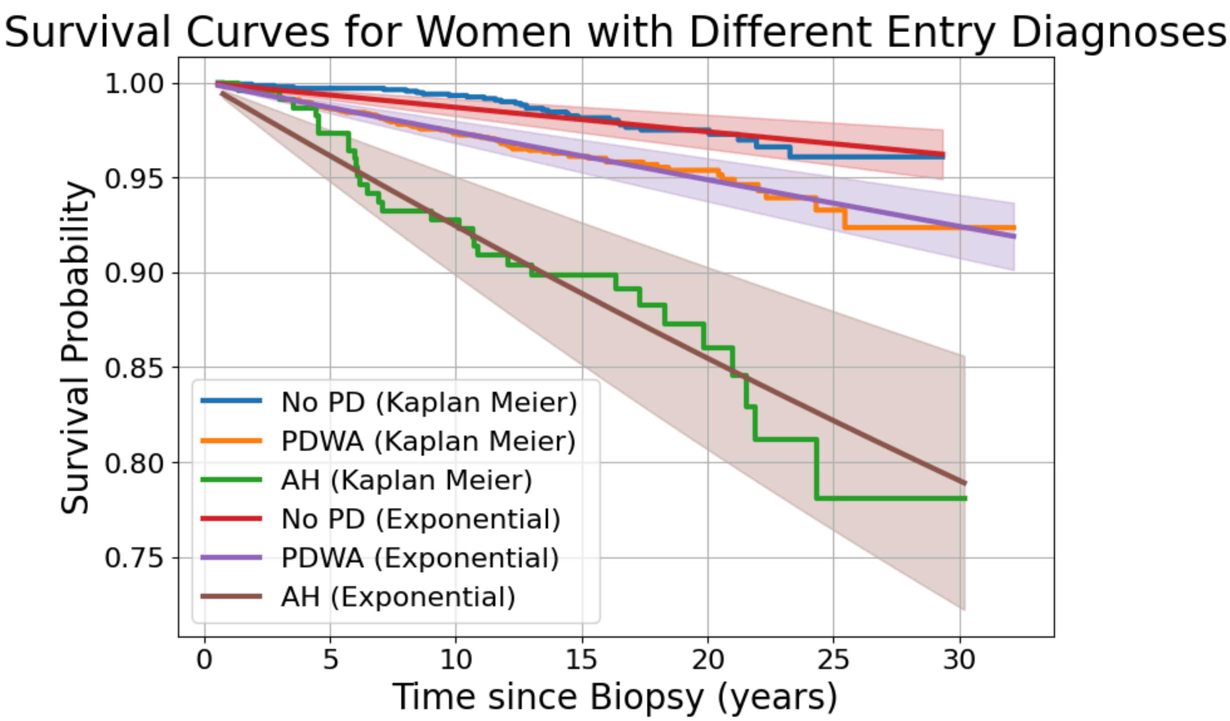
(multiply by -1)

$$\underbrace{-\frac{d}{dt} \log S_1(t)}_{H_1(t)} = \underbrace{-R\frac{d}{dt} \log S_0(t)}_{H_0(t)}$$

$$H_1(t) = R\, H_0(t)$$

$$\therefore \frac{H_1(t)}{H_0(t)} = R \quad \& \quad \frac{h_1(t)}{h_0(t)} = R$$

Exercise 4



Exponential Distribution Estimates: The shaded areas represent the estimated survival curves for each group using the exponential distribution. The exponential distribution assumes a constant hazard rate over time, leading to a straight-line survival curve. Here, the exponential estimates appear as straight lines with varying slopes for each group.

Interpretation:

- The survival curves estimated using the exponential distribution do not perfectly align with the Kaplan-Meier estimates. This indicates that the assumption of a constant hazard rate inherent in the exponential distribution may not accurately capture the true survival patterns observed in the data.
- Specifically, the curves for AH and PDWA groups show notable deviations from the exponential distribution assumptions, with more complex shapes over time. This suggests that the hazard rates for these groups may not be constant but instead vary over time.
- One example interpretation for Kaplan Meier would be that for AH it is 90% likely to survive longer than 15 years

- The No PD group's survival curve appears closer to the exponential distribution assumption, indicating that the constant hazard rate assumption might be more plausible for this group. However, there are still some deviations, especially at later time points, suggesting that the exponential distribution might not be the best fit for this group either.

Code for Reference

```python
from lifelines import KaplanMeierFitter, ExponentialFitter
plt.figure(figsize=(15, 7))
fig, ax = plt.subplots()
exponential_fitter = ExponentialFitter()
group_name_dict = {
  0: "No PD",
  1: "PDWA",
  2: "AH"
}
for group_name, group_data in breast_data.groupby("pd"):
    time, prob_survival = kaplan_meier_estimator(
        group_data.fate == 1, group_data.follow
    )

    ax.step(time, prob_survival, where="post", label=group_name_dict[group_name] + " (Kaplan Meier) ")
for group_name, group_data in breast_data.groupby("pd"):
    exponential_fitter.fit(group_data.follow, event_observed=group_data.fate == 1, label=group_name_dict[group_name] + " (Exponential) ")
    exponential_fitter.plot_survival_function()

plt.title('Survival Curves for Women with Different Entry Diagnoses')
plt.xlabel('Time since Biopsy (years)')
plt.ylabel('Survival Probability')
plt.grid(True)
ax.legend(loc="best")
```