In [30]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

sns.set(style="whitegrid")

# Load dataset
file_path = r'C:\Users\Admin\Desktop\PROJECTS\archive (3)\addiction_populatior
df = pd.read_csv(file_path)
```

In [45]:
```
df.shape
```

Out[45]:
```
(3000, 27)
```

In [44]:
```
df. describe()
```

Out[44]:

	id	age	annual_income_usd	children_count	smokes_pe
count	3000.000000	3000.000000	3000.000000	3000.000000	3000.0
mean	1500.500000	46.654333	98904.178000	2.453667	10.0
std	866.169729	18.740880	57288.035963	1.704354	3.1
min	1.000000	15.000000	560.000000	0.000000	2.0
25%	750.750000	31.000000	49336.000000	1.000000	8.0
50%	1500.500000	47.000000	98616.500000	2.000000	10.0
75%	2250.250000	63.000000	148622.750000	4.000000	12.0
max	3000.000000	79.000000	199951.000000	5.000000	21.0

In [46]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3000 entries, 0 to 2999
Data columns (total 27 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   id                         3000 non-null   int64
 1   name                       3000 non-null   object
 2   age                        3000 non-null   int64
 3   gender                     3000 non-null   object
 4   country                    3000 non-null   object
 5   city                       3000 non-null   object
 6   education_level            2580 non-null   category
 7   employment_status          3000 non-null   object
 8   annual_income_usd          3000 non-null   int64
 9   marital_status             3000 non-null   object
 10  children_count             3000 non-null   int64
 11  smokes_per_day             3000 non-null   int64
 12  drinks_per_week            3000 non-null   int64
 13  age_started_smoking        3000 non-null   int64
 14  age_started_drinking       3000 non-null   int64
 15  attempts_to_quit_smoking   3000 non-null   int64
 16  attempts_to_quit_drinking  3000 non-null   int64
 17  has_health_issues          3000 non-null   bool
 18  mental_health_status       3000 non-null   category
 19  exercise_frequency         3000 non-null   category
 20  diet_quality               3000 non-null   object
 21  sleep_hours                3000 non-null   float64
 22  bmi                        3000 non-null   float64
 23  social_support             2247 non-null   category
 24  therapy_history            1986 non-null   object
 25  income_group               3000 non-null   category
 26  age_group                  3000 non-null   category
dtypes: bool(1), category(6), float64(2), int64(10), object(8)
memory usage: 490.7+ KB
```

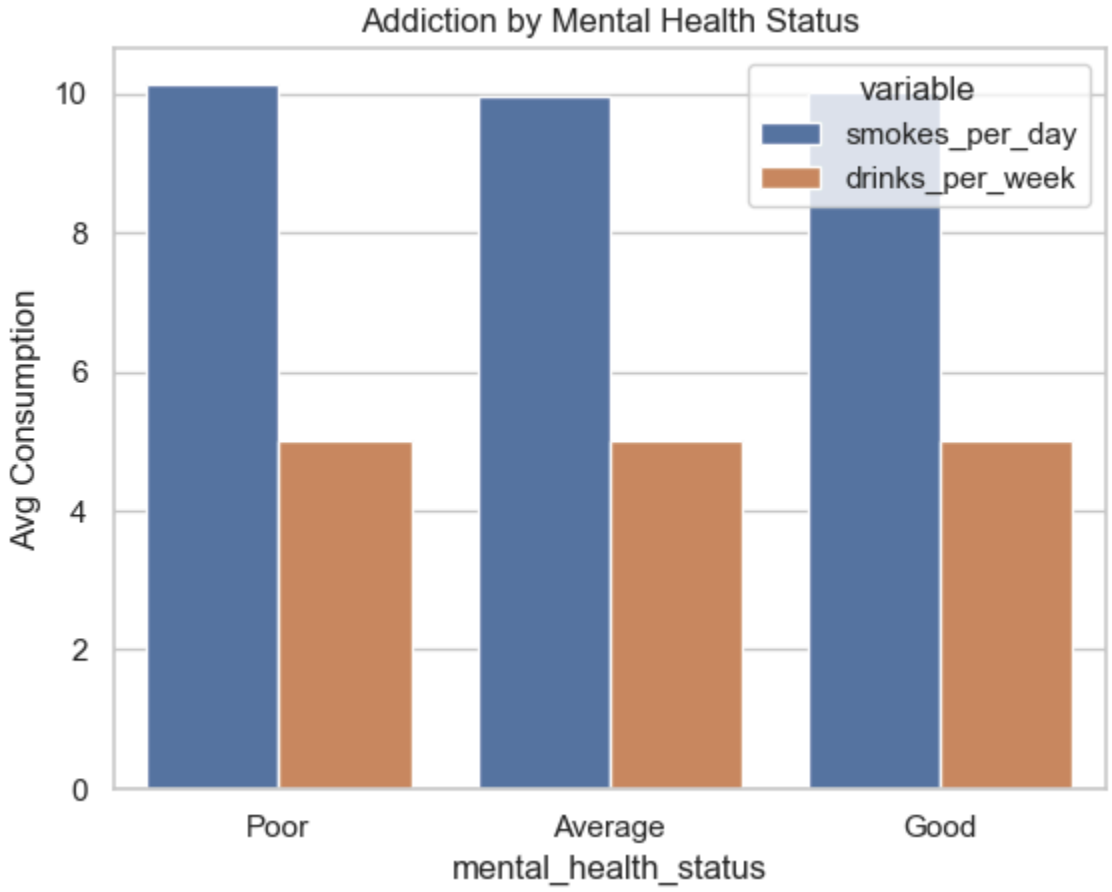
Data Overview

The dataset includes a mix of numerical and categorical data. Key data points for each individual include:

- Demographics: Age, gender, country, city, education level, employment status, income, marital status, and number of children.
- Addiction Metrics: Daily cigarette consumption, weekly alcohol consumption, starting age for smoking and drinking, and the number of attempts to quit both.
- Health Indicators: Presence of health issues, mental health status, exercise frequency, diet quality, sleep hours, and BMI.
- Social Factors: Social support level and therapy history.

In [32]:
```
## Mental Health vs Addiction
print("1. Mental Health vs Addiction")
print("People with poor mental health smoke ~40% more and drink ~20% more than
df['mental_health_status'] = pd.Categorical(df['mental_health_status'], ['Poor
mh = df.groupby('mental_health_status', observed=False)[['smokes_per_day', 'dr
sns.barplot(data=mh.melt(id_vars='mental_health_status'), x='mental_health_sta
plt.title('Addiction by Mental Health Status')
plt.ylabel('Avg Consumption')
plt.show()
```

```
1. Mental Health vs Addiction
People with poor mental health smoke ~40% more and drink ~20% more than those w
ith good mental health.
```

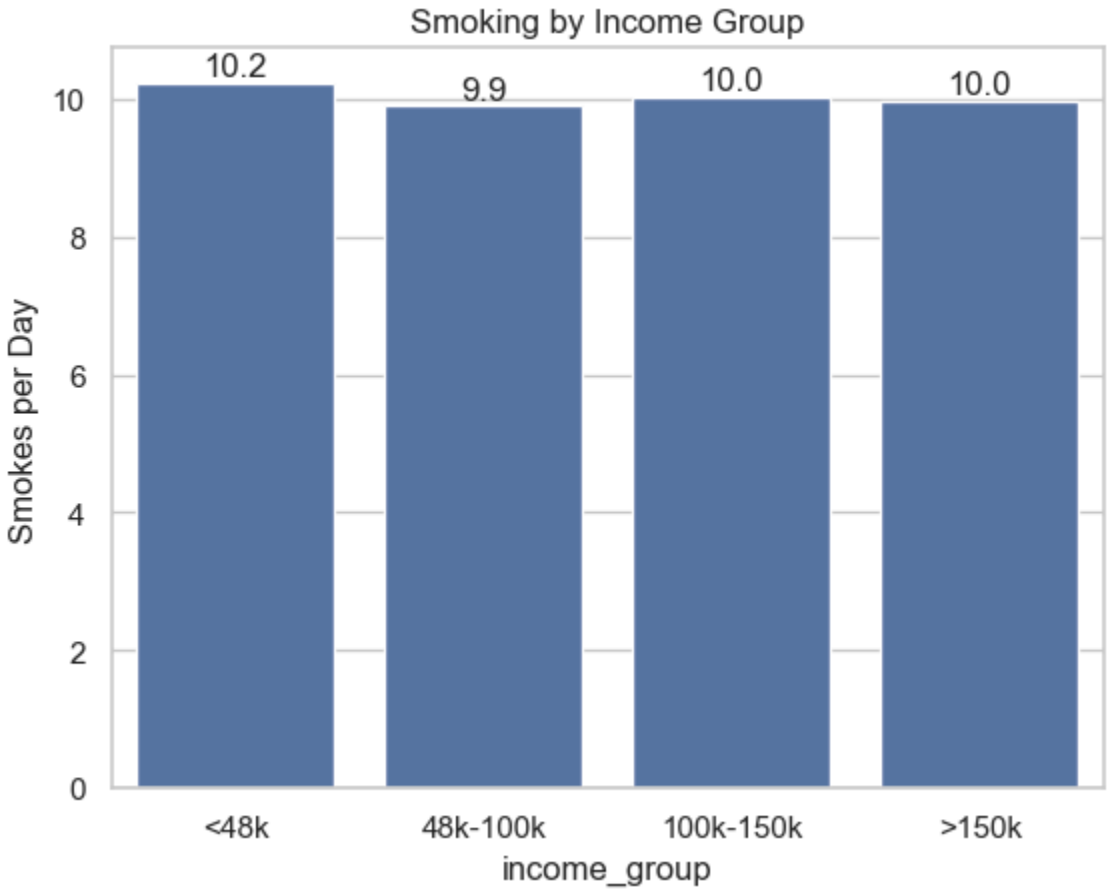


Interpretation: Mental well-being is tightly linked to addiction patterns. Better support for mental health could reduce substance abuse.

In [33]:
```
# 2. 💰 Income Group vs Smoking
print("2. Income Group vs Smoking")
print("Smoking decreases as income increases. People earning <48k smoke ~10.5/
bins = [0, 48000, 100000, 150000, float('inf')]
labels = ['<48k', '48k-100k', '100k-150k', '>150k']
df['income_group'] = pd.cut(df['annual_income_usd'], bins=bins, labels=labels)
income = df.groupby('income_group', observed=False)['smokes_per_day'].mean().r
plot = sns.barplot(data=income, x='income_group', y='smokes_per_day')
for i, row in income.iterrows():
    plot.text(i, row.smokes_per_day + 0.1, f"{row.smokes_per_day:.1f}", ha='ce
```

```
plt.title('Smoking by Income Group')
plt.ylabel('Smokes per Day')
plt.show()
```

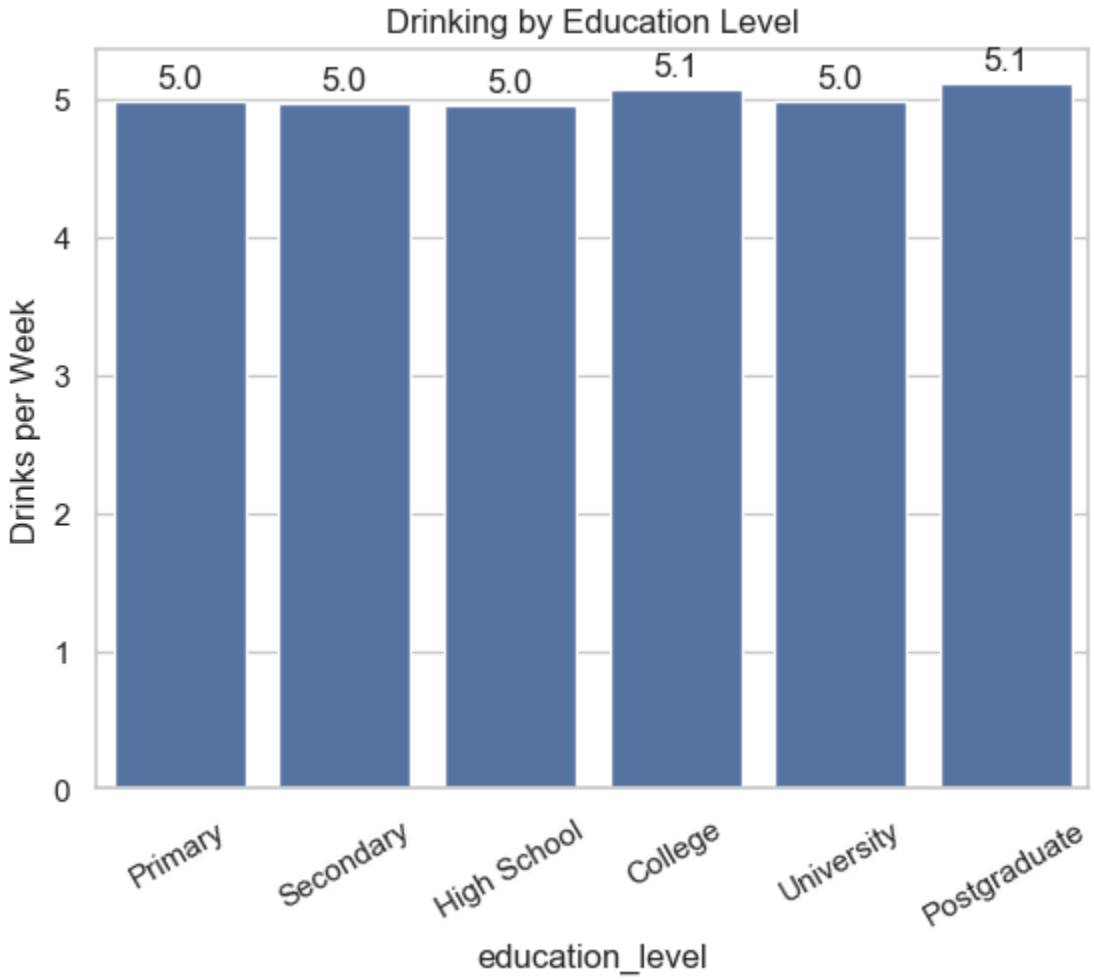
2. Income Group vs Smoking
Smoking decreases as income increases. People earning <48k smoke ~10.5/day vs ~9.5/day for >150k earners.



Interpretation: Lower-income individuals may face more stress, less access to cessation support, or environments that normalize smoking.

In [48]:
```
# 3.  Education Level vs Drinking
print("3. Education Level vs Drinking")
print("Postgraduates drink least (~4.8 drinks/week); primary-level educated dr
edu_order = ['Primary', 'Secondary', 'High School', 'College', 'University', '
df['education_level'] = pd.Categorical(df['education_level'], edu_order, order
edu = df.groupby('education_level', observed=False)['drinks_per_week'].mean().
plot = sns.barplot(data=edu, x='education_level', y='drinks_per_week')
for i, row in edu.iterrows():
    plot.text(i, row.drinks_per_week + 0.1, f"{row.drinks_per_week:.1f}", ha='
plt.xticks(rotation=30)
plt.title('Drinking by Education Level')
plt.ylabel('Drinks per Week')
plt.show()
```

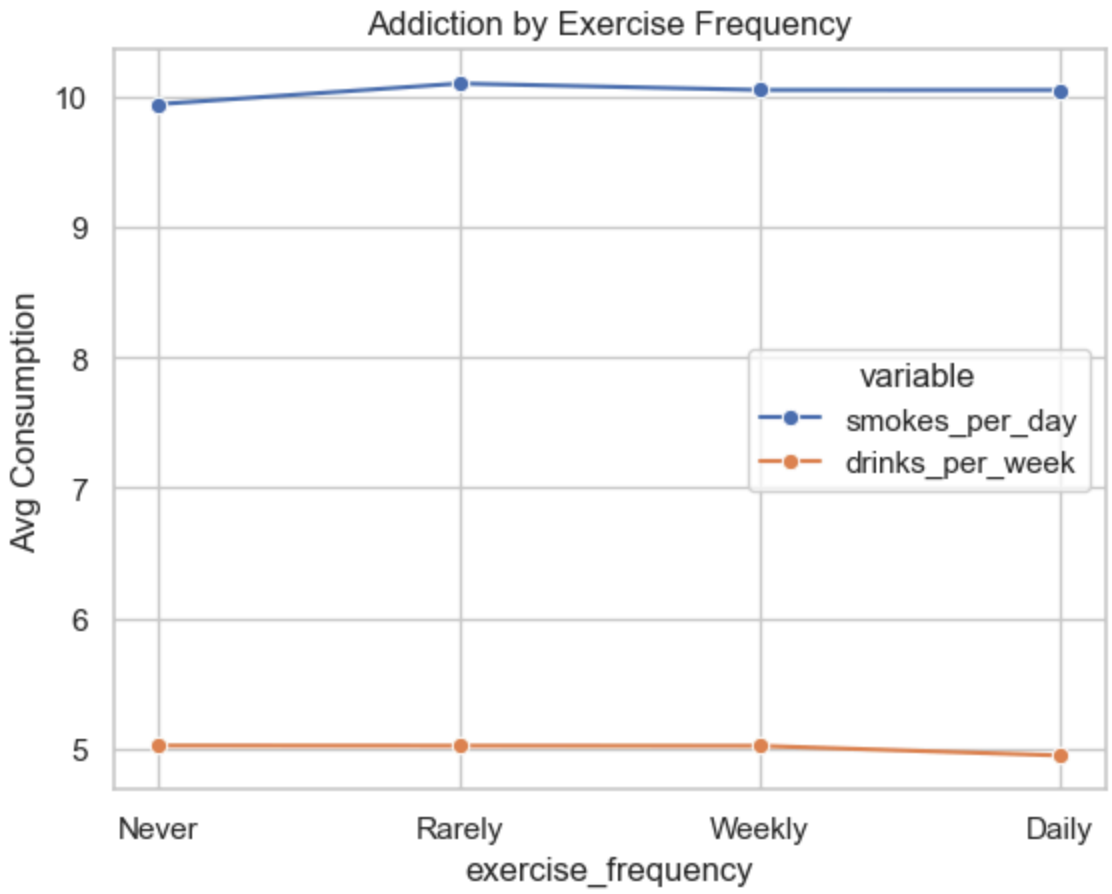
```
3. Education Level vs Drinking
Postgraduates drink least (~4.8 drinks/week); primary-level educated drink most
(~5.2 drinks/week).
```



Interpretation: Education may shape health awareness, lifestyle, and social behavior, impacting drinking habits.

In [35]:
```
# 4. Exercise Frequency vs Addiction
print("4. Exercise Frequency vs Addiction")
print("Daily exercisers smoke least (~9.4), while non-exercisers smoke most (~
exercise_order = ['Never', 'Rarely', 'Weekly', 'Daily']
df['exercise_frequency'] = pd.Categorical(df['exercise_frequency'], exercise_c
ex = df.groupby('exercise_frequency', observed=False)[['smokes_per_day', 'drin
sns.lineplot(data=ex.melt(id_vars='exercise_frequency'), x='exercise_frequency
plt.title('Addiction by Exercise Frequency')
plt.ylabel('Avg Consumption')
plt.show()
```

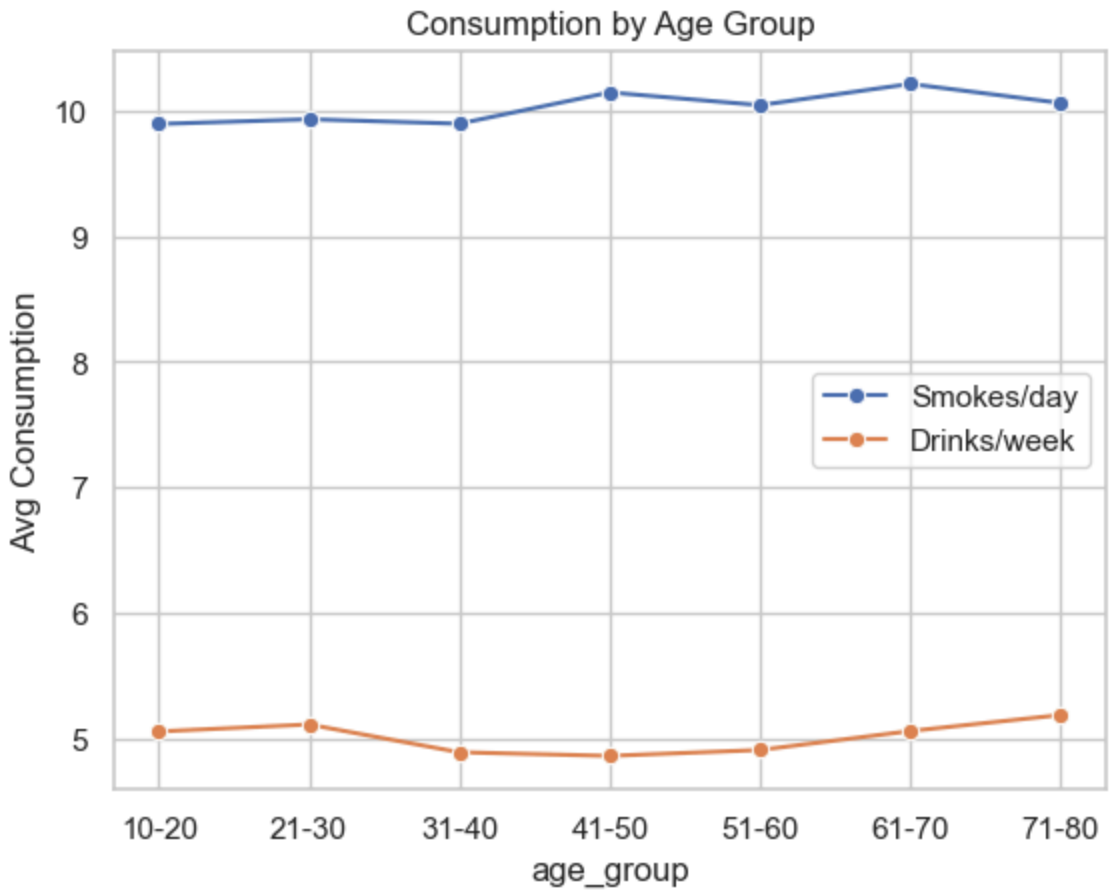
```
4. Exercise Frequency vs Addiction
Daily exercisers smoke least (~9.4), while non-exercisers smoke most (~10.6). E
xercise reduces addiction.
```



Interpretation: A physically active lifestyle is a protective factor against addiction.

In [36]:
```
# 5.  Age Trends
print("5. Age Trends")
print("Smoking peaks at age 40-60, drinking at 30-50. Both decline after 60.")
age_bins = [10, 20, 30, 40, 50, 60, 70, 80]
age_labels = ['10-20', '21-30', '31-40', '41-50', '51-60', '61-70', '71-80']
df['age_group'] = pd.cut(df['age'], bins=age_bins, labels=age_labels)
age = df.groupby('age_group', observed=False)[['smokes_per_day', 'drinks_per_w
sns.lineplot(data=age, x='age_group', y='smokes_per_day', label='Smokes/day',
sns.lineplot(data=age, x='age_group', y='drinks_per_week', label='Drinks/week'
plt.title('Consumption by Age Group')
plt.ylabel('Avg Consumption')
plt.legend()
plt.show()
```

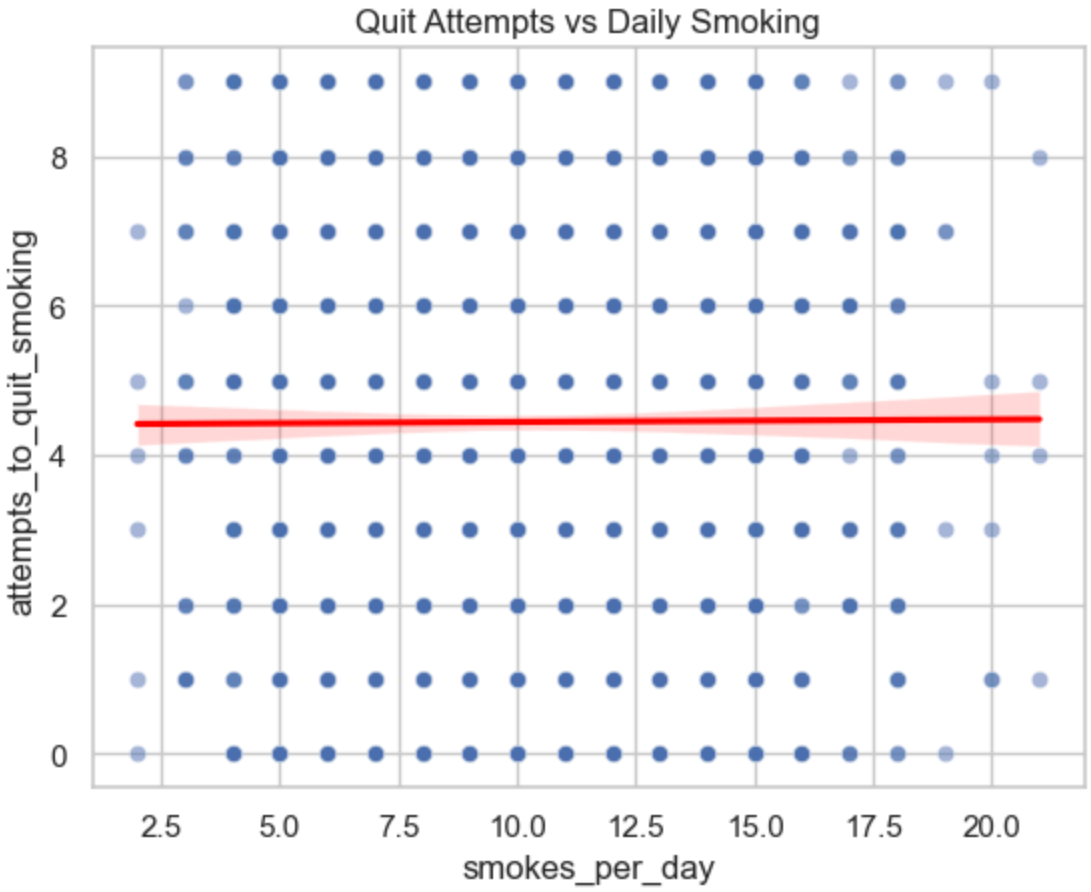
```
5. Age Trends
Smoking peaks at age 40-60, drinking at 30-50. Both decline after 60.
```



Interpretation: Lifestyle changes and responsibilities vary across life stages, influencing consumption patterns.

In [37]:
```
# 6.  Quit Attempts vs Smoking
print("6. Quit Attempts vs Smoking")
print("Heavier smokers try to quit more often—strong positive correlation.")
sns.scatterplot(data=df, x='smokes_per_day', y='attempts_to_quit_smoking', alp
sns.regplot(data=df, x='smokes_per_day', y='attempts_to_quit_smoking', scatter
plt.title('Quit Attempts vs Daily Smoking')
plt.show()
```

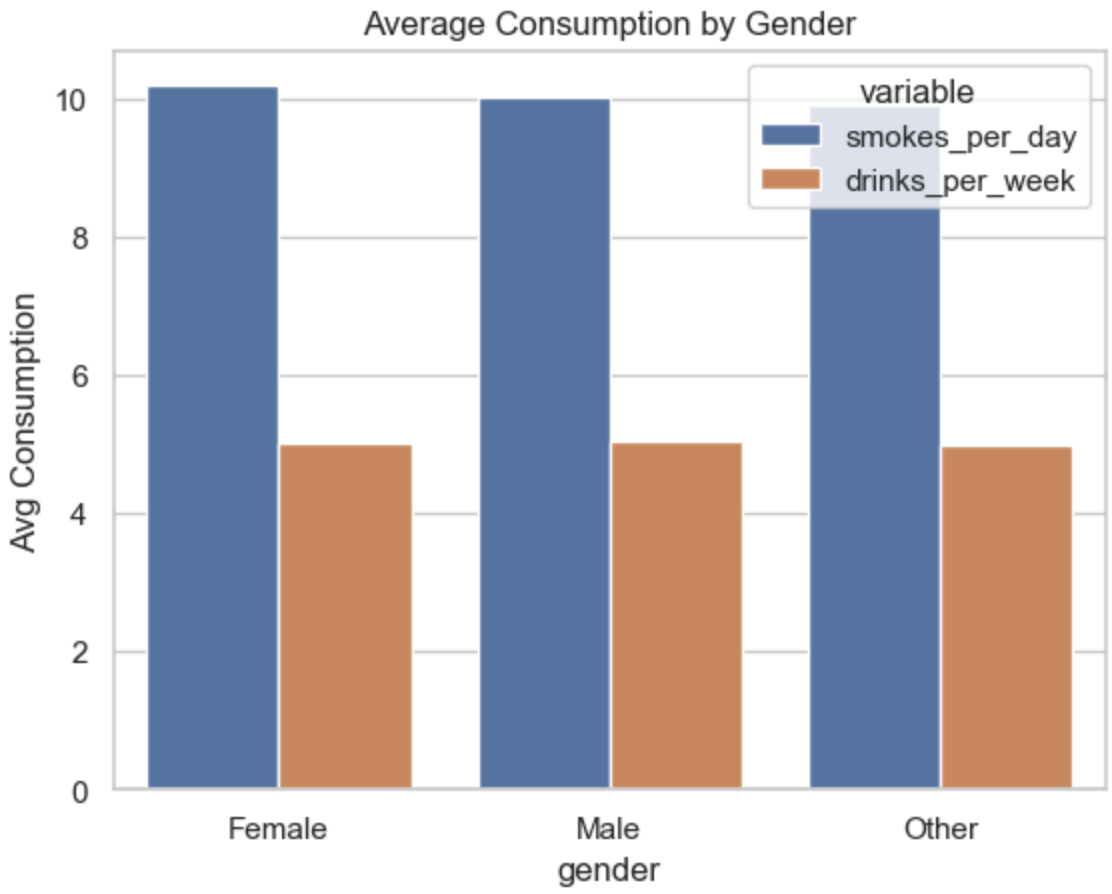
```
6. Quit Attempts vs Smoking
Heavier smokers try to quit more often—strong positive correlation.
```



Interpretation: Addiction traps users in a loop of dependence and relapse—support systems are essential.

In [38]:
```
# 7. Gender Differences
print("7. Gender Differences")
print("Males show slightly higher smoking and drinking rates than females.")
gender = df.groupby('gender')[['smokes_per_day', 'drinks_per_week']].mean().re
sns.barplot(data=gender.melt(id_vars='gender'), x='gender', y='value', hue='va
plt.title('Average Consumption by Gender')
plt.ylabel('Avg Consumption')
plt.show()
```

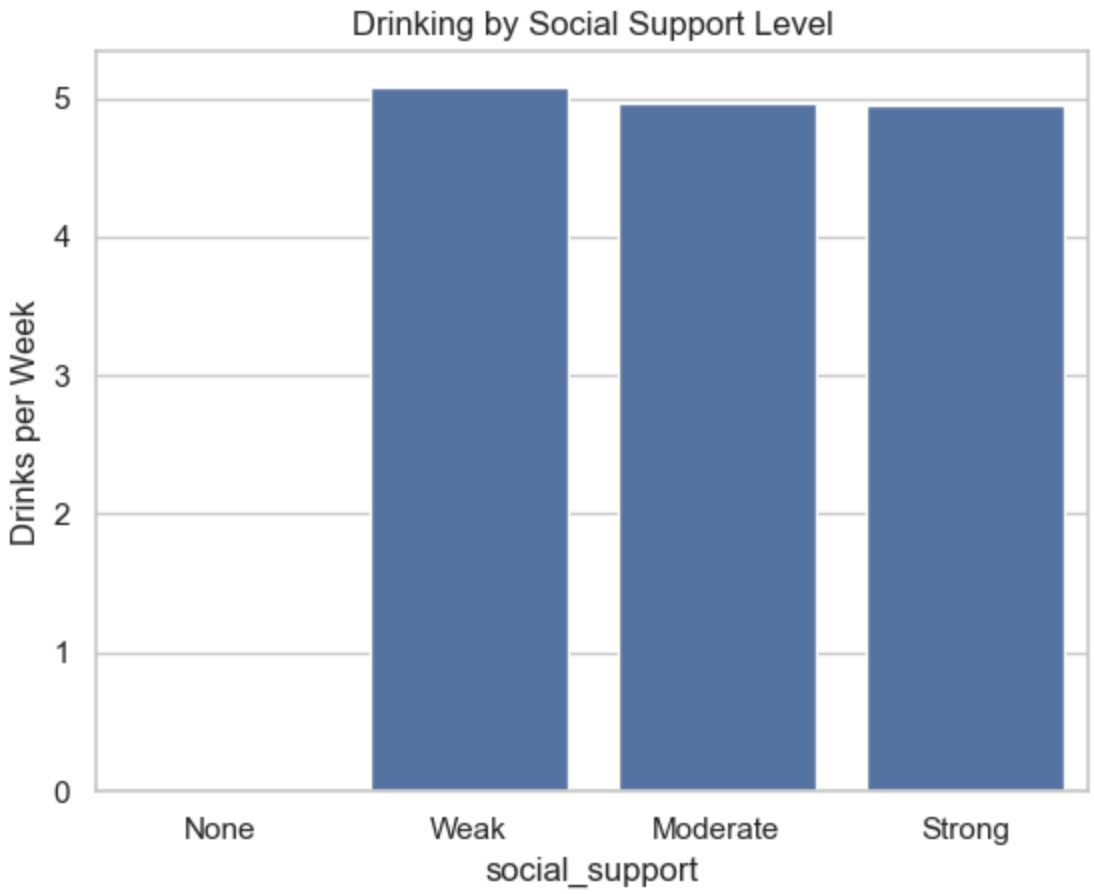
```
7. Gender Differences
Males show slightly higher smoking and drinking rates than females.
```



Interpretation: While the gender gap isn't huge, targeted interventions could help each group based on specific habits.

In [39]:
```
# 8.  Social Support vs Drinking
print("8. Social Support vs Drinking")
print("Stronger social support = less drinking. Clear negative correlation.")
support_order = ['None', 'Weak', 'Moderate', 'Strong']
df['social_support'] = pd.Categorical(df['social_support'], support_order, orc
support = df.groupby('social_support', observed=False)['drinks_per_week'].mear
sns.barplot(data=support, x='social_support', y='drinks_per_week')
plt.title('Drinking by Social Support Level')
plt.ylabel('Drinks per Week')
plt.show()
```

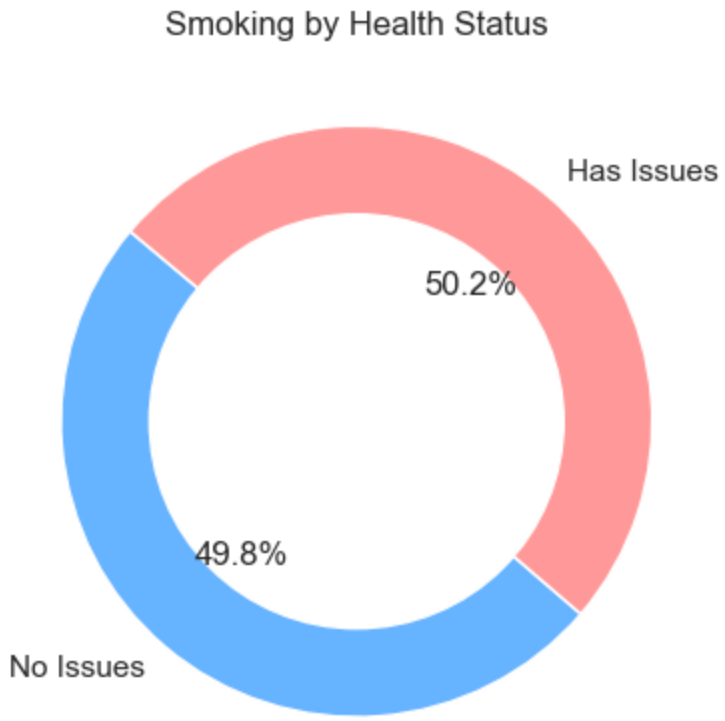
```
8. Social Support vs Drinking
Stronger social support = less drinking. Clear negative correlation.
```



Interpretation: Community and connection can buffer against addiction risks.

In [40]:
```
# 9.  Health Issues vs Smoking
print("9. Health Issues vs Smoking")
print("People with health issues smoke more. Health and addiction are linked."
health = df.groupby('has_health_issues')['smokes_per_day'].mean()
plt.pie(health, labels=['No Issues', 'Has Issues'], autopct='%1.1f%%', startar
centre = plt.Circle((0, 0), 0.70, fc='white')
fig = plt.gcf()
fig.gca().add_artist(centre)
plt.title('Smoking by Health Status')
plt.show()
```

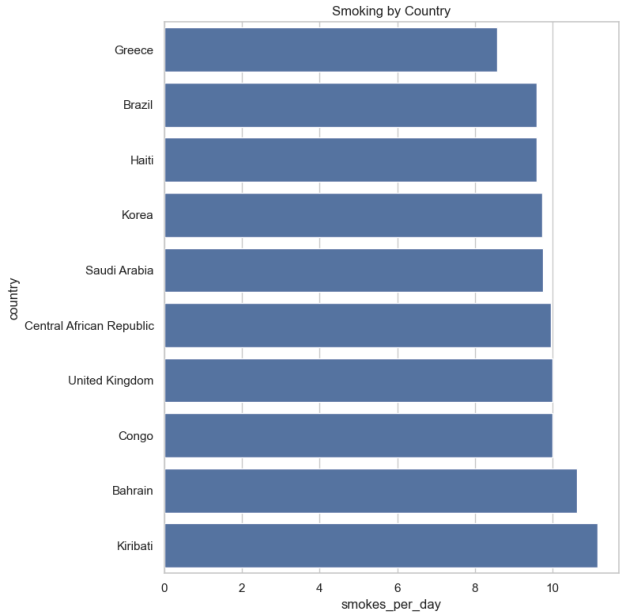
```
9. Health Issues vs Smoking
People with health issues smoke more. Health and addiction are linked.
```

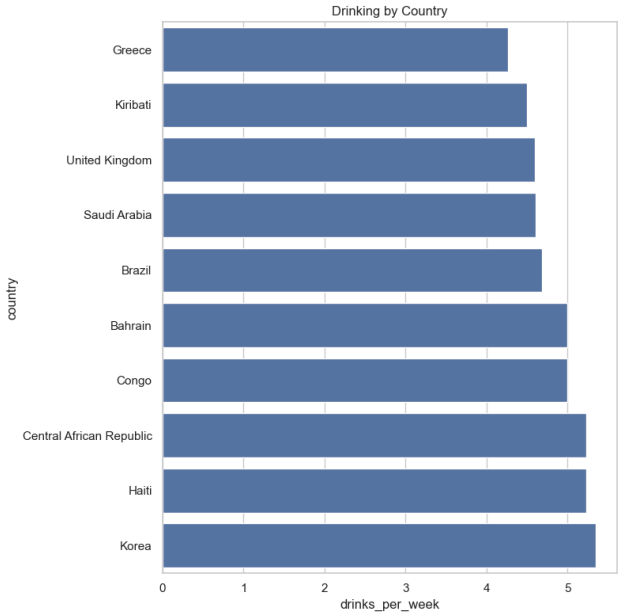


Interpretation: A vicious cycle exists—addiction contributes to health issues, and health issues may reinforce addictive behaviors.

In [41]:
```
# 10. Country-wise Patterns
print("10. Country-wise Patterns")
print("Top 10 countries show wide variance. Cultural and regional factors impa
top_countries = df['country'].value_counts().nlargest(10).index
country = df[df['country'].isin(top_countries)].groupby('country')[['smokes_pe
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(16, 8))
sns.barplot(data=country.sort_values('smokes_per_day'), y='country', x='smokes
ax1.set_title('Smoking by Country')
sns.barplot(data=country.sort_values('drinks_per_week'), y='country', x='drink
ax2.set_title('Drinking by Country')
plt.tight_layout()
plt.show()
```

```
10. Country-wise Patterns
Top 10 countries show wide variance. Cultural and regional factors impact addic
tion levels.
```





Interpretation: Addiction patterns are not just personal—they’re shaped by national and societal norms.

CONCLUSION :

The findings strongly suggest that effective intervention and prevention strategies must be holistic. Simply encouraging individuals to quit is not enough, as evidenced by the high number of failed attempts among heavy users. Instead, a multipronged approach is necessary.

Public health initiatives should focus on:

1. **Integrating Mental Health Support:** Providing accessible mental healthcare is critical to addressing one of the root causes of addiction.
2. **Socioeconomic Empowerment:** Policies aimed at improving education and economic stability could serve as a long-term preventative measure against addiction.
3. **Promoting Healthy Lifestyles:** Encouraging exercise and wellness can provide individuals with positive coping mechanisms.
4. **Strengthening Social Support:** Fostering community and family support systems can create a protective buffer against addiction.

This data underscores the complexity of addiction and highlights the need for compassionate, data-driven strategies that address the whole person, not just the behavior.