Gamma GLM: Exam Reference Sheet

1. Interpreting Gamma GLM Output

- **Family:** Gamma with log link models positive, skewed severity.
- **Intercept:** Baseline mean severity on log scale. exp(Intercept) = base severity.
- **Coefficients:**
  - Estimate > 0: increases mean severity; < 0: decreases it.
  - exp(coef) = multiplicative effect on mean severity.
  - E.g., coef = 0.2 $\Rightarrow$ exp(0.2) $\approx$ 1.22: ˜22% higher severity.
  - Check significance: p-value < 0.05 = statistically significant.
- **Residual Deviance:** Lower = better fit. Compare to null deviance.
- **Dispersion:** Should be close to 1. If high, consider overdispersion.
- **Deviance Residuals:** Should center around 0. Outliers = poor fit.

2. Model Fit Metrics

- **AIC**: Lower AIC = better trade-off between fit and complexity.
- **BIC**: Penalizes complexity more than AIC. Prefer smaller BIC for large n.
- **Null vs Residual Deviance**: Large drop = model explains variation well.

3. Prediction Error Metrics

- **MSE:** Mean of squared errors. Penalizes large errors. Targets mean.
- **RMSE:** $\sqrt{\text{MSE}}$, interpretable in outcome units (e.g. dollars).
- **MAE:** Mean absolute error. Targets median. Less sensitive to outliers.
- **GI Pricing Insight:** Use MSE/RMSE to target mean severity. MAE can mislead due to skew MAE focuses on Median, which is less than Mean in skewed data, we would under reserve.
- **R code:**

```
pred <- predict(model, newdata=test, type="response")
mean((pred - test$y)^2)      # MSE
sqrt(mean((pred - test$y)^2)) # RMSE
mean(abs(pred - test$y))     # MAE
```

4. Residual Diagnostics

- **Residuals vs Fitted:** Random scatter $\Rightarrow$ good. Pattern $\Rightarrow$ misspecification.

- **Q-Q Plot:** Deviance residuals ≈ normal. Tail deviation = poor fit.
- **Scale-Location:** Checks variance homogeneity. Flat line desired.
- **Cook's Distance:** Identifies influential observations. Large values = outliers.
- **Good Fit:** Residuals centered, Q-Q linear, no big Cook's points.

5. Model Pros and Improvements

Pros

- Log link + Gamma handles skewed severity.
- Coefficients interpretable on percentage scale.
- Lower AIC/BIC than null = better fit.
- Predicts mean cost: aligns with pricing needs.

Improvements

- Remove non-significant predictors.
- Add interaction terms (e.g. Age × Gender).
- Use splines/bands for nonlinear effects.
- Check outliers: cap or treat separately.
- Consider alt. distributions (Log-normal, Tweedie) if fit poor.
- Validate on hold-out data. Compare RMSE.