Assessment: Regression and Baseball, part 1

Question 1

1/1 point (graded)

What is the final linear model (in the video "Building a Better Offensive Metric for Baseball") we used to predict runs scored per game?

- `lm(R ~ BB + HR)`
- `lm(HR ~ BB + singles + doubles + triples)`
- `lm(R ~ BB + singles + doubles + triples + HR)` ✔
- `lm(R ~ singles + doubles + triples + HR)`

Answer

Correct: Correct.

Explanation

`lm(R ~ BB + singles + doubles + triples + HR)` is the only one of the models above that predicts runs scored based on all of the following: BBs, singles, doubles, triples, and HRs.

Submit

You have used 1 of 2 attempts

Answers are displayed within the problem

Question 2

1/1 point (graded)

We want to estimate runs per game scored by individual players, not just by teams. What summary metric do we calculate to help estimate this?

Look at the code from the video "Building a Metter Offensive Metric for Baseball" for a hint:

```
pa_per_game <- Batting %>%
  filter(yearID == 2002) %>%
  group_by(teamID) %>%
  summarize(pa_per_game = sum(AB+BB)/max(G)) %>%
  .$pa_per_game %>%
  mean
```

The summary metric used is:

- ○ `pa_per_game`: the mean number of plate appearances per team per game for each team
- ○ `pa_per_game`: the mean number of plate appearances per game for each player
- ● `pa_per_game`: the number of plate appearances per team per game, averaged across all teams ✔

Explanation

`pa_per_game` is the number of plate appearances per team per game averaged across all teams. We initially calculated the `pa_per_game` grouped by teams but then took the means across all teams to get one summary metric.

Submit You have used 1 of 1 attempt

ℹ Answers are displayed within the problem

Question 3

1/1 point (graded)

Imagine you have two teams. Team A is comprised of batters who, on average, get two bases on balls, four singles, one double, and one home run. Team B is comprised of batters who, on average, get one base on balls, six singles, two doubles, and one triple.

Which team scores more runs, as predicted by our model?

- ○ Team A
- ● Team B
- ○ Tie

- Impossible to know

✔

Answer
Correct: Correct.

Explanation
By using the coefficients from the linear model to predict the number of runs scored by each team, you find that Team B is expected to score more runs on average.

Submit You have used 1 of 2 attempts

ℹ Answers are displayed within the problem

Question 4

1/1 point (graded)
The on-base-percentage plus slugging percentage (OPS) metric gives the most weight to:

- Singles
- Doubles
- Triples
- Home Runs

✔

Answer
Correct: Correct.

Explanation
By looking at the equation for OPS, you can tell that the OPS metric weights home runs most heavily.

$$\text{OPS} = \frac{\text{BB}}{\text{PA}} + \frac{\text{Singles} + 2\text{Doubles} + 3\text{Triples} + 4\text{HR}}{\text{AB}}$$

Submit You have used 1 of 2 attempts

ℹ Answers are displayed within the problem

Question 5

1/1 point (graded)
What statistical concept properly explains the “sophomore slump”?

- (•) Regression to the mean
- ( ) Law of averages
- ( ) Normal distribution

✔

Explanation
Regression to the mean is what explains the sophomore slump. The correlation for performance in two separate years is high but not perfect, so high performers will tend to perform slightly worse in the following year (and low performers will tend to perform slightly better in the following year).

Submit You have used 1 of 1 attempt

ℹ Answers are displayed within the problem

Question 6

1/1 point (graded)
In our model of time vs. observed_distance in the video "Measurement Error Models", the randomness of our data was due to:

- ( ) sampling
- ( ) natural variability
- (•) measurement error

✔

Explanation
Measurement error models look at applications where randomness is introduced from measurement error instead of sampling or natural variability.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 7

1/1 point (graded)

Which of the following are important assumptions about the measurement errors in the experiment presented in the video "Measurement Error Models"?

Select ALL that apply.

- [x] The measurement error is random
- [x] The measurement error is independent
- [x] The measurement error has the same distribution for each time i

✔

Explanation

In this model, we asumed that the measurement errors were random, independent from each other, and had the same distribution for each time i.

We also assumed that there was no bias, which means that $\mathrm{E}\left[\varepsilon\right] = 0$.

Submit

You have used 1 of 2 attempts

Answers are displayed within the problem

Question 8

1/1 point (graded)

Which of the following scenarios would violate an assumption of our measurement error model?

- [ ] The experiment was conducted on the moon.
- [x] There was one position where it was particularly difficult to see the dropped ball.
- [ ] The experiment was only repeated 10 times, not 100 times.

✔

Explanation

If there were one position where it was particularly difficult to see the dropped ball, that would violate the assumption of randomness. If the experiment were conducted on the moon, that would simply predict a different gravitational constant. Repeating the experiment 10 instead of 100 times would not matter because we do not need a large sample for our assumptions to be valid in this model.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem